Modelling dispersal in plants

Modellierung der Ausbreitung bei Pflanzen

Dissertation zur Erlangung des Doktorgrades

der Naturwissenschaften (Dr. rer. nat.)

dem Fachbereich Biologie

der Philipps-Universität Marburg

vorgelegt von

Marc Niggemann

aus Arnsberg

Marburg/Lahn

Juli 2009

Vom Fachbereich Biologie der Philipps-Universität Marburg
als Dissertation am ……. 2009 angenommen.

Erstgutachter: Prof. Dr. Birigt Ziegenhagen
Zweitgutachter: PD Dr. Volker Grimm

Tag der mündlichen Prüfung ……2009

One never notices
what has been done;
one can only see
what remains to be done.

(Marie Curie)

Contents

Erklärung: Eigene Beiträge und veröffentlichte Teile der Arbeit

Der §8, Absatz 3 der Promotionsordnung der naturwissenschaftlichen Fachbereiche der Philipps-Universität Marburg in der Fassung vom 12. April 2000 schreibt vor, dass „die individuellen Leistungen des Doktoranden deutlich abgrenzbar und bewertbar sein müssen", falls Teile der Dissertation aus gemeinsamer Forschungsarbeit entstanden. Dies betrifft die Publikationen und Manuskripte 1 – 3 (im folgenden Paper 1 – 3) sowie die Anhänge A – C (im folgenden Appendix A – C). Die individuellen Beiträge werden im Folgenden näher erläutert.

Paper 1 Distribution patterns of plants explained by human movement behaviour.

- Ausarbeitung des Modells und Erstellung des Simulationsprogramms
- Durchführung und Auswertung aller Simulationen
- Verfassen des Manuskripts in Zusammenarbeit mit Dr. S. Brunzel, Dr. S. Jetzkowitz, Dr. M.C. Wichmann und Dr. R. Bialozyt
- Veröffentlichung: Niggemann, M., Brunzel, S., Jetzkowitz, J., Wichmann, M.C., Bialozyt, R. (2009). Ecological Modelling, 220: 1339-1346.
  Das vorliegende Kapitel entspricht der veröffentlichten Version.

Paper 2 Marked point pattern analysis on genetic paternity data for estimation and uncertainty assessment of pollen dispersal kernels

- Durchführung der Vaterschaftsanalysen
- Auswertung der Ergebnisse in Zusammenarbeit mit Dr. T. Wiegand und Dr. J.J. Robledo-Arnuncio
- Verfassen des Manuskripts in Zusammenarbeit mit Dr. T. Wiegand, Dr. J.J. Robledo-Arnuncio und Dr. R. Bialozyt
- Das vorliegende Kapitel entspricht der eingereichten Version.

Paper 3 Pollen competition and restricted gene flow between *Populus nigra* and its hybrid form *Populus x canadensis*

- Ausarbeitung und Erstellung des Simulationsprogramms
- Durchführung und Auswertung aller Simulationen
- Verfassen des Manuskripts

Appendix A Conceptual work for a new wind dispersal kernel

- Ausarbeitung des Modells in Zusammenarbeit mit Dr. R. Bialoyzt
- Aufstellen der Gleichungssysteme in Zusammenarbeit mit Dr. J. Vollmer
- Durchführung der Berechnungen
- Verfassen des Manuskripts

Appendix B Flower phenology of *Populus nigra* and *Populus x canadensis*

- Beobachtungen im Gelände in Zusammenarbeit mit G. Rathmacher
- Auswertung der Beobachtungen und Vergleich mit der Literatur
- Verfassen des Manuskripts

Appendix C Germination experiments of *Populus nigra* and *Populus x canadensis*

- Ernte und Ansaat der Samen in Zusammenarbeit mit G. Rathmacher
- Auswertung der Ergebnisse
- Verfassen des Manuskripts

Über diese Arbeiten hinaus habe ich während meiner Doktorarbeit im Rahmen meines Aufenthaltes am Centre for Ecology and Hydrology in Dorset (UK) unter der Leitung von Dr. Matthias C. Wichmann und Dr. James Bullock an einem thematisch verwandten Projekt mitgearbeitet, das nicht Teil meiner Dissertation ist. Diese Arbeit befasste sich mit der Ausbreitung von Pflanzensamen über die Schuhe von Menschen. Sowohl bei den Experimenten als auch bei der Ausarbeitung der zugrunde liegenden Ausbreitungsfunktion im Modell war ich beteiligt. Die Arbeit ist bei Proceedings of the Royal Society of London B veröffentlicht worden.

- Wichmann, M.C., Alexander, M.J., Soons, M.B., Galsworthy, S., Dunne, L., Gould, R., Fairfax, C., Niggemann, M., Hails, R.S., Bullock, J.M., 2009. Human-mediated dispersal of seeds over long distances. Proceedings of the Royal Society B. 276, 523-532.

Außerdem hatte ich Anteil an der Entstehung der Bachelorarbeit von Sebastian Flinkerbusch mit dem Thema „Simulation der Bewegungsmuster von Krallenaffen (Callithrichidae) im Kontext der Samenausbreitung von *Parkia panurensis*“ (2008). Sowohl bei der Entstehung des Modells als auch bei der Umsetzung in ein Simulationsprogramm und der abschließenden Auswertung war ich beteiligt.

Des Weiteren habe ich zusammen mit Dr. Ronald Bialozyt die Diplomarbeit von Melanie Köhnen mit dem Thema „Die Wirkung verschiedener reproduktiver Hintergründe in einem natürlichen *Populus nigra – P. x canadensis* Hybridkomplex auf Performance-Parameter von Einzelbaumabsaaten“ (2008) betreut. Dabei wirkte ich bei der Entwicklung und Durchführung der ökologischen Experimente als auch bei der Diskussion der Ergebnisse mit.

Die Beteiligung an den gemeinsamen Veröffentlichungen bzw. Manuskripten mit G. Rathmacher erstrecken sich auf die Probennahme, die Zusammenstellung und Aufarbeitung der genetischen und räumlichen Daten mittels GIS, den Vaterschaftsanalysen der geernteten Samen sowie die Diskussion der Ergebnisse und redaktionelle Teile der Manuskripte.

- Rathmacher, G., Niggemann, M., Wypukol, H., Gebhardt, K., Ziegenhagen, B., Bialozyt, R. (2009). Allelic ladders and reference genotypes for a rigorous standardization of poplar microsatellite data. Trees: Structure and Function 23, 573-583.

- Rathmacher, G., Niggemann, M., Köhnen, M., Ziegenhagen, B., Bialozyt, R. Short-distance gene flow in *Populus nigra* L. accounts for small-scale spatial genetic structures - implications for in-situ conservation measures. Conservation Genetics, accepted.

Die Anfertigung der Dissertation in englischer Sprache wurde vom Dekan des Fachbereiches Biologie, Prof. Dr. ……., am …….. 2009 genehmigt.

SYNTHESIS

1. General introduction

Dispersal is the key process in plant species population dynamics. The importance of the process has been pointed out in several main publications (Ridley 1930, “The dispersal of plants throughout the world”; Harper 1977, “Population biology of plants”). Dispersal enables plants as static organisms to move. With this ability plants are able to escape unfriendly conditions, exist in a changing environment and expand their distribution range.

On the individual level, following the Janzen-Connell hypothesis, plant offspring has to avoid intraspecific competition primarily by the mother plant (Janzen 1970). Thus, seeds have to be dispersed out of the mother’s range of influence. On the population level, there is the urgent need for connections and therefore for an exchange of individuals between different habitat sites, especially to re-establish populations where they go extinct from time to time, based on the concept of metapopulations (Levins 1970, “Extinction”).

Besides these fundamental ecological theories, conservationists also want to know the consequences for species with naturally limited or nowadays restricted dispersal abilities in our rapidly changing landscape. For some species with only short dispersal distances ongoing fragmentation of the landscape leads to isolated populations threatened by extinction. These populations will be lost because of environmental changes, stochastic events or genetic effects.

Gene flow

Gene flow is the transfer of alleles between populations by exchange of individuals, seeds or pollen. If gene flow from neighbouring populations is sufficient, small populations suffer less from local inbreeding depression because local loss of alleles could be compensated by the metapopulation. Otherwise, genetic variation between those populations is reduced and local

adaptation or even speciation processes are prevented. Interspecies gene flow leads to hybridisation processes. The offspring of these crossings are able to show enhanced traits compared to their parents. If these hybrids successfully and regularly cross back with one of their parents, we found introgressive gene flow and the formation of a hybrid swarm.

Dispersal mechanisms

Coming back to the ecological basics of dispersal there are different dispersal mechanisms to connect populations and to reach distant habitat sites by pollen and seeds. Some plants take an active part by catapulting their seeds away from themselves. Others rely in abiotic processes or biotic agents. Dispersal of seeds by water flow respectively rivers (=hydrochory) and by wind (=anemochory) are the most common abiotic processes leading to the spread of plants by seeds and even by vegetative plant material. There are several kinds of animal dispersal mechanisms. Seeds are transported by these biotic agents either on their skin (=epizoochory) or within their digestive tract until they are excreted (=endozoochory). However, the dispersal of pollen is limited to drifting by wind or transport by animals, in temperate regions mainly by bumblebees and in tropical regions by bats and birds.

Most important for the connection and colonisation of suitable habitats are long distance dispersal (LDD) events. Following Reid's paradox these events are necessary to explain today's plant distribution, because mean dispersal distances of most trees and herbs are too small to account for recolonisation of northern latitudes after the last ice age (Reid 1899). However, LDD is not common but very rare. Therefore, it is hardly measured because sample size to detect even these rare events is always limited. LDD is possible by wind especially strong updrafts (Nathan et al. 2002) but also by animal dispersal. Seeds have been found several hundred kilometres away from the source carried by sheep (Manzano and Malo 2006) or birds (Sorensen 1986).

Interestingly, there are species with typical dispersal mechanisms for short distances, e.g. by rhizomes, but today they are dispersed over long distances and become very frequent as aliens in the new colonised regions. This makes it clear that species with seemingly typical mechanisms for a special kind of dispersal like hooks for epizoochorus animal dispersal or umbrellas for wind dispersal are not restricted to their typical vectors. Specialised dispersal vectors may be most suitable but common vectors are more frequent. Therefore we have to keep in mind that morphological characteristics do not exclude other dispersal vectors. Especially under human influence combinations and changes in the dispersal vector are possible. This makes it complicated to understand the process that leads to the observed species distribution.

Simulation models

To analyse these biological processes simulation models become very popular tools. They allow forecasts and risk scenarios as well as nearly unlimited repetitions under the same circumstances what is not possible in experiments. Therefore models are used as an addition to experiments and field data. However, this is not their main advantage to which they are often reduced. Models make it possible to go one step further and reveal the important processes behind observations that cannot be identified without extensive experiments and field work. Simulation models often aim to answer general questions or reproduce the overall pattern instead of giving exact values or picturing exactly the whole world. To do this sufficiently models should be preferred that follow the statement of Albert Einstein: “Things should be made as simple as possible, but not simpler”. Models should focus on a distinguished question to really make a reduction of parameters to the basic components of the system possible.

Ecological relations form the underlying processes that are responsible for the resulting pattern. To reveal these processes models are created that are able to reproduce the pattern examined in our natural system. This is called pattern oriented modelling (POM) where the simulation of the natural pattern gives evidence of the underlying processes (Wiegand 2003). The examined pattern could be the distribution of a species in the landscape, a time of species extinction or other eye catching observations. The advantage of this approach is that a model is able to find the most relevant processes for these patterns and to simulate scenarios under different conditions e.g. climate change.

Within the category of ecological simulation models there are differential equation models (Wissel 1989) as well as individual-based models (Grimm and Railsback 2005). In contrast to differential equation models that are based on clear mathematical formulas and typically do not include random processes and individual behaviour, rule based models like individual-based models (IBMs) have the opportunity to easily translate ecological relations into model assumptions and rules. This makes it possible to directly introduce expert knowledge into a model. The fundamental advantage of IBMs, however, is the individual-based approach. The actual internal and external environments of a separate unit lead to individual behaviour that in a second step forms the condition and reaction of the whole population. This bottom up approach of adaptive behaviour creating higher organisation levels follows the observations we find in natural systems.

OUTLINE OF THIS DISSERTATION

In this dissertation, I present two simulation models focussing on different dispersal types in plants. The first model is dealing with human mediated dispersal of ruderal plants in a cultural landscape (Paper 1) whereas the second model looks into details of pollen mediated gene flow in the wind pollinated tree species *Populus nigra* in a natural stand. Basic ideas of modelling wind dispersal lead to the concept of a new wind dispersal kernel (Appendix A). The necessary components and parameters for the dispersal model of our study species were received from phenological observations (Appendix B), from germination experiments (Appendix C) and from a kernel fitting to the results of a spatial point pattern analysis (Paper 2). The implementation of the model made it possible to conduct simulation studies giving further insights in the restriction of the crossing between *P. nigra* and its hybrid form *P. x canadensis* (Paper 3).

2. Human mediated dispersal

In historical times rural landscapes with changing habitats and stock farming were dominated by moving flocks of cattle and sheep that act as dispersal vectors for many plants. But today, this animal mediated dispersal is often reduced and replaced by human mediated dispersal, because on the one hand modern stock farming became more static and on the other hand human mobility and traffic increased dramatically during the last 20 years. Since the end of the 18th century it was already proposed that humans have a huge impact on the dispersal of plants (Willdenow 1798; Humboldt and Bonpland 1807, Ridley 1930). There have been attempts to quantify (Bullock et al. 2003, von der Lippe and Kowarik 2008, Wichmann et al. 2009) or at least give qualitative evidence for the human influence (Salisbury 1942, Clifford 1956, Hodkinson and Thompson 1997, von der Lippe and Kowarik 2007). However, it only could be shown that seeds in principle are able to be dispersed by human vectors, but the consequences for plant distribution are still unknown. To overcome these limitations the combination of sociological and ecological data has been realised as a necessary goal to understand the system dynamics in the human sphere of influence. Buchan and Padilla (1999) compared successfully the use of motorboats in waterways in the US state of Wisconsin with the invasion of the zebra mussel *Dreissena polymorpha*. Human movement behaviour was a better predictor of species movement and especially long distance dispersal events than pure distance dependent diffusion models.

The same pattern that human movement behaviour affects species dispersal was found for different ruderal plant species in settlements in Central Germany (Brunzel et al. 2005). The change in the distribution of these species over 20 years could be explained by the amount of human movement between the settlements in the studied region (Paper 1). Interestingly, the strong effect of human movement behaviour on plant distribution was predominantly shown by alien species. Those expand their range using the distance independent movement patterns

of humans who create connections between sites because of the movement to their job, their relatives and friends and because of the shopping infrastructure of the settlements. In the fate of increased human mobility the importance of humans as dispersal vectors will be constantly increasing.

Human behaviour as driver for dispersal can be used indirectly by transferring human movement between locations into a static dispersal matrix (see Appendix Paper 1) or directly as rules in individual-based models. We successfully applied the approach of modelling individual movement behaviour for seed dispersal on nonhuman primates and the distribution of seeds of their food plant *Parkia panurensis* (Flinkerbusch 2008).

3. Wind dispersal

There is a common dispersal pattern for almost all plant species of the temperate zone regarding the pollen and seed shadow. With increasing distance to the mother plant the density of dispersed pollen and seeds decreased more or less exponentially, whereas the slope and the tail of the dispersal curve are more species specific. The simplest assumption is a distribution based on statistical probability density functions like Gaussian or exponential functions (Austerlitz et al. 2004). The problem is that these functions are in principle one dimensional and show an even distribution for all directions. However, this is only true for some landscapes. For example, if wind directions are limited to some orientations because of landscape elements, the estimation of the pollen or seed shadow will be misleading using circular functions. Therefore, meteorological components are important to incorporate wind conditions like direction, strength and the overall wind stability of the atmosphere (Appendix A). The problem is that the incorporation of these parameters into simple and widely used distribution functions has not been done sufficiently and complex meteorological models require input parameters that are not recorded regularly but need substantial equipment and time for data collection.

Another controversial discussed topic especially in wind dispersal is the tail of the dispersal curve. The tail, which is responsible for long distance dispersal events, has no fixed starting point. Furthermore, it is possible that after a first dispersal event a second or even third event takes place (secondary / tertiary dispersal) masking the tail of the curve. A typical example is the dispersal first by catapulting the seeds away from the mother plant and secondary by ants feeding on an elaiosomes attached to the seed. This combination provides fine scale dispersal into suitable habitats by the ants. Additionally, occasional long distance dispersal events are able to produce different results compared to the usually observed seed shadow, but these events are only rare and often remain unseen but are able to rewrite the common dispersal

pattern. Therefore the tail of the dispersal curve responsible for long distance dispersal events strongly depends on the number of samples in these distances. Consequently, the use and estimation of a dispersal kernel should be elaborated and according to the scientific question.

3.1 Dispersal kernel estimation

There are different ways to receive a suitable dispersal kernel for a given population or species. A common approach is the fitting of observed dispersal distances of seeds from their mothers to a probability density function. The collection of these distances by direct observation is hardly possible for seeds and impossible for the dispersal of pollen. Today with ongoing improvement in genetic methods genetic paternity analyses become the technique of choice to determine the dispersal distance of pollen and seeds (Sork and Smouse 2006). The analyses are based on genetic fingerprints of the offspring to identify the father respectively mother plants. Microsatellites are used to address each plant individually. This makes it not only possible to carry out genetic paternity analyses but to conduct analyses on genetic diversity, introgressive gene flow and spatial genetic structure of the studied population to even reveal dispersal processes of the past.

A statistical approach to determine the dispersal distances is inverse modelling. Out of a random sample of seeds the gene flow from potential parents is estimated without genetic paternity analyses. If the total number of pollen and seeds a species is able to disperse is well known, this method is a powerful tool to easily describe gene flow by seeds (Sagnard et al. 2007).

However, this and other techniques do not account for the spatial geometry of a population and the distribution of fathers and mothers. Statistical confidence in the paternity data is not considered, especially at long distances, where we can expect low signal-to-noise ratios. Spatial point pattern fulfils this requirement and, what is even more important, is able to give

a range where we can rely on our data points from the genetic paternity analysis (Paper 2). Up to now this range of reliance is totally ignored in previous studies with the consequence of a misleading view of gene flow especially for long distances. In our study we found effective fathers in a distance up to 7,500m within a huge sample size of 1,535 seeds, but we could only reveal a significant difference from random shuffling of father trees up to 2,000m and a distance dependency up to 310m. Furthermore a two component exponential power function could be fitted to the data of the point pattern analysis. Because this new kernel function was similar to a comparable kernel fitted to our data using the competing sources model of Robledo-Arnuncio and Garcia (2007), our spatial statistics methods of marked point pattern analysis was not only suitable to calculate the range where our data provides significant distance dependency but could also provide a suitable dispersal kernel for pollen dispersal. The fitted kernel can be used for further simulations regarding our studied population.

3.2 Introgression risk in *Populus nigra*

Black poplars (*Populus nigra* L.) are typical trees of the softwood forest vegetation along rivers. The species is dioecious and wind pollinated. Each individual is able to produce up to 10 Billion pollen grains or 50 Million seeds per season (OECD 2000). Since the introduction of the American Black Poplar (*Populus deltoides*) in Europe in the 17th century a hybrid (*Populus x canadensis*) with the European Black Poplar becomes very frequent in the landscape. Because of the excellent growth rate and habit this hybrid form is used in the timber industry. Additionally, it has a straight habit and no overhanging branches like the natural *P. nigra* habit, so that it has been planted as ornamental trees in Western European countries to picture the Mediterranean flair and to serve as windbreak between fields. Backcrosses with their parents and therefore introgressive gene flow is possible (Vanden Broeck et al. 2004, Ziegenhagen et al. 2008) leading to the risk of intermixed genotypes with

P. nigra that is already threatened because of habitat loss and missing natural dynamic along the more and more embanked rivers.

The flower phenology seems to act as a natural barrier to inhibit reproduction and gene flow between the two species. The hybrid *P. x canadensis* is usually flowering earlier than *P. nigra*. However, depending on climate conditions, the start of flowering varies and the previously distinct flowering times are able to overlap (Appendix B). Especially after a cold and long winter and a spring with fast rising temperatures the flowering will start in both species nearly at the same time. This is an irregular phenomenon but may become common in future climate change scenarios. The consequence is the formation of a hybrid swarm crossing back and forth with the natural species. This introgressive gene flow softens the pure species status of *P. nigra*.

Another barrier is the incompatibility of hybrid pollen on *P. nigra* females. Hand pollination experiments have shown that hybrid pollen is less effective than *P. nigra* pollen (Benetka et al. 2002; Vanden Broeck et al. 2003). This would explain the surprising low introgression rates in natural stands but the effectiveness was not tested directly in open pollinated trees.

This disadvantage of hybrid pollen in a real landscape could be quantitatively shown by using dispersal kernels to simulate dispersal in a natural mixed stand and then comparing the proportion of incoming hybrid pollen and the genetically confirmed introgression rates (Paper 3). The modelling approach made it possible to unravel values for the reproductive barrier between *P. nigra* and *P. x canadensis* in the same range as hand pollination experiments have concluded. If the location and the species status of male poplar trees are known, this approach will be suitable to estimate introgression rates in other regions, too.

However, it has to be mentioned, that introgression rates resulting from small seed samples can be misleading because of stochastic effects. The high hybrid disadvantage to become effective of up to a level of 10^{-2} is able to reduce the detectable hybrid influence in the sampled seeds to zero apparently, but introgressive gene flow still occurs even though in low

numbers. Although hybrids seem to be less effective than the natural black poplar trees, the risk of introgression is only reduced and not impossible. Furthermore, we have to bear in mind, that gene flow is also possible the other way round by *P. nigra* pollen becoming effective on female hybrids where we do not found any barrier restricting gene flow between *P. nigra* and *P. x canadensis*. Additionally, seeds and offspring resulting from crossings between the two species are vital, but we could show that the germination rates of hybrid mothers are lower than in *P. nigra* (see Appendix C). However, after this initial stage the seedlings of hybrid mothers perform even better and seem to become better competitors than seedlings of *P. nigra* mothers during the establishment in different dense grass levels (Köhnen 2008). Whatever crossing direction is successful, if seeds are produced, there is a high possibility of the establishment of introgressed offspring leading to a hybrid swarm in the *P. nigra – P. x canadensis* complex.

In addition to the loss of the pure species status of *P. nigra*, another thread is already at hand affecting not only this single species. The next step in the breeding forest institutions is the introduction of genetic modifications into hybrid poplars for further increase of the timber production. These genetic modified trees (GM trees) will improve wood qualities and will be more resistant against herbivorous insects. The consequences for the outnumbered natural *P. nigra* as well as for the whole system of floodplain forests are not foreseeable.

4. Conclusion

Interdisciplinary research and new methods

Models usually require expert knowledge from various scientific and non-scientific fields. The ecological models presented in this thesis resulted from cooperation with botanists, foresters, geneticists, meteorologists, physicists, sociologists and statisticians.

In the first model the strong effect of human movement behaviour on plant distribution especially of alien species could be shown. Therewith sociological studies should be integrated into questions concerning the spread of alien plants in human dominated landscapes, because these species are often the beneficiary of increased human mobility. The same as we are looking at animal dispersers we should account for humans as the new and increasing community of dispersers.

In the second model, working on a new wind dispersal kernel, the necessity to introduce meteorological insights into wind dispersal processes became essential. Other disciplines not only have a differentiated view on the same things, but sometimes they already have answers. New methods and techniques are helpful to discover so far unseen processes. Applying genetic paternity analyses on ecological questions is such a technique able to detect even rare dispersal events. Subsequently, using the sophisticated spatial statistic approach of point pattern analysis the common underestimation of uncertainty in kernel estimation could be revealed.

This interdisciplinarity is the key to overcome the limitations and restrictions of each discipline alone. Creating models of combined data sets may be the first step. However, it has to be mentioned that although interdisciplinarity is demanded from science foundations and societies as well as from scientific journals, it is not really rewarded, because there is still separation between disciplines based on lacking overview and acceptance.

Kernel estimation

Although the introduction of a suitable wind dispersal kernel seems to be necessary for modelling the pollen dispersal of a wind pollinated species like *P. nigra*, the results using a modified non-circular diffusion model showed no better explanation than standard kernels. As

long as density distribution functions are suitable for fitting experimental data with kernel functions, there seem to be no need for wind dependent functions.
The use of spatial statistic and marked point pattern analyses made it possible to fit a dispersal kernel to our data set of fathered seeds for poplar mother trees. However, the short range of significant deviation from the null model of random shuffling of father trees shows the considerable limitations of the use of fitted kernel functions, even when the fit based on a huge data set of sampled and analysed seeds.

Modelling approach

Modelling plant dispersal between settlements using humans as dispersal vectors was a successful implementation. The model could not only demonstrate the effect of human behaviour on plant species distribution but it was able to differentiate between the modelled species regarding their effect on human movement and gardening styles.
The integration of the fitted kernel into a simulation model for pollen dispersal in *P. nigra* showed the possibility to calculate the barrier for hybrid pollen pollinating female poplars in open pollinated trees. This clarifies the importance of a modelling approach. The recalculation of the potential amount of pollen reaching a female tree to compare the proportion of hybrid fathers in the potential pollen cloud with the actual proportion in the sampled seed set can be done for other populations as well. However, the species status and the distance to the mother have to be known for potential father trees.
With a suitable pollen dispersal kernel at hand, the evidence for a barrier for hybrid pollen, the knowledge of flower phenology, germination and establishment, several simulation studies become possible. A future scenario would be the development of the natural *P. nigra* stand in the following years under the impact of increased hybrid pollen influx and the aging of the population as well as the creation of new sites by different water fluctuations.
After all, the successful combination of modelling techniques with sociological data and genetic parentage analyses made it possible to go one step further and to better understand the processes of plant dispersal and the resulting pattern of plant distribution in our landscape. Especially the fact that even simple models are able to explain huge parts of observed effects should encourage the modelling community to go on.

5. References

Austerlitz, F., Dick, C.W., Dutech, C., Klein, E.K., Oddou-Muratorio, S., Smouse, P.E., and Sork, V.L. (2004). Using genetic markers to estimate the pollen dispersal curve. Molecular Ecology 13, 937-954.

Benetka, V., Vackova, K., Bartakova, I., Pospiskova, M. and Rasl, M. (2002). Introgression in black poplar (*Populus nigra* L. ssp. *P. nigra*) and its transmission. Journal of Forest Science 48, 115-120.

Brunzel, S., Fischer, S.F., Jetzkowitz, J. and Schneider, J. (2005). Die Zusammenhänge zwischen den Veränderungen in der dörflichen Ruderalflora und dem Wandel ländlicher Lebensstile. Schlussbericht des Forschungsprojektes. VW-Stiftung, Hannover.

Buchan, L.A.J. and Padilla, D.K. (1999) Estimating the probability of long-distance overland dispersal of invading aquatic species. Ecological Applications 9, 254–265.

Bullock, J.M., Moy, I.L., Coulson, S.J. and Clarke, R.T. (2003) Habitat-specific dispersal: environmental effects on the mechanisms and patterns of seed movement in a grassland herb *Rhinanthus minor*. Ecography 26, 692–704.

Clifford, H.T. (1956) Seed dispersal on footwear. In: Proceedings of the Botanical Society of the British Isles 2, 129–131.

Fiedler, T. (2005) Modellierung der Interaktion von Lebensstiltypen und Ruderalflora. Diploma thesis in Jena, Germany.

Flinkerbusch, S. (2008) Simulation der Bewegungsmuster von Krallenaffen (*Callithrichidae*) im Kontext der Samenausbreitung von *Parkia panurensis*. Bachelor thesis in Marburg, Germany

Grimm, V. and Railsback, S.F. (2005) Individual-based Modeling and Ecology. Princeton University Press, Princeton, New Jersey.

Harper, J.L. (1977) Population biology of plants. London. Academic Press.

Hodkinson, D.J. and Thompson, K. (1997) Plant dispersal: the role of man. Journal of Applied Ecology 34, 1484–1496.

Humboldt, A.v. and Bonpland, A. (1807) Ideen zu einer Geographie der Pflanzen. Nachdruck 1963, Darmstadt. Wiss. Buchgesellschaft.

Janzen, D.H. (1970) Herbivores and the number of tree species in tropical forests. The American Naturalist 104, 501-528.

Köhnen, M. (2008) Die Wirkung verschiedener reproduktiver Hintergründe in einem natürlichen Populus nigra – P. x canadensis Hybridkomplex auf Performance-Parameter von Einzelbaumabsaaten. Diploma thesis in Marburg, Germany.

Levins, R. (1970) Extinction. In: Gerstenhaber, M. (ed.): Some mathematical problems in biology. Providence, R.I.: American Mathematical Society, 77-107.

Manzano, P. and Malo, J.E. (2006) Extreme long-distance seed dispersal via sheep. Frontiers in Ecology and the Environment 4, 244-248.

Nathan, R., Katul, G.G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W., and Levin, S.A. (2002) Mechanisms of long-distance dispersal of seeds by wind. Nature, 418, 409-413.

OECD. Consensus document on the biology of Populus L. (poplars), Organisation for Economic Co-operation and Development, Series on Harmonization of Regulatory Oversight in Biotechnology No. 16, Paris, 2000.

Reid, C. (1899) The origin of the British Flora. London: Dulau.

Ridley, H.N. (1930) The dispersal of plants throughout the world. Ashford, UK. Reeve & Co.

Robledo-Arnuncio, J.J. and Garcia, C. (2007) Estimation of the seed dispersal kernel from exact identification of source plants. Molecular Ecology 16, 5098-5109.

Sagnard, F., Pichot, C., Dreyfus, P., Jordano, P. and Fady, B. (2007) Modelling seed dispersal to predict seedling recruitment: Recolonization dynamics in a plantation forest. Ecological Modelling 203, 464-474.

Salisbury, E.J. (1942) The weed problem. Nature 149, 594–596.

Sorensen, A.E. (1986) Seed dispersal by adhesion. Annual Review of Ecology and Sytematics 17, 443-463.

Sork, V.L., and Smouse, P.E. (2006) Genetic analysis of landscape connectivity in tree populations. Landscape Ecology 21, 821-836.

Vanden Broeck, A., Quataert, P., Roldán-Ruiz, I., Van Bockstaele, E., and J. Van Slycken (2003). Pollen competition in *P. nigra* females revealed by microsatellite markers. Journal of Forest Genetics. 10, 219–227.

Vanden Broeck, A., Storme, V., Cottrell, J.E., Boerjan, W., Van Bockstaele, E., Quataert, P., and Van Slycken, J. (2004) Gene flow between cultivated poplars and native black poplar (*Populus nigra* L.): a case study along the river Meuse on the Dutch-Belgian border. Forest Ecology and Management 197, 307-310.

von der Lippe, M. and Kowarik, I. (2007) Long-distance dispersal of plants by vehicles as a river of plant invasions. Conservation Biology 4, 986–996.

von der Lippe, M. and Kowarik, I. (2008) Do cities export biodiversity? Traffic as dispersal vector across urban–rural gradients. Diversity and Distributions 14, 18–25.

Wichmann, M.C., Alexander, M.J., Soons, M.B., Galsworthy, S., Dunne, L., Gould, R., Fairfax, C., Niggemann, M., Hails, R.S. and Bullock, J.M. (2009) Human-mediated dispersal of seeds over long distances. Proceedings of the Royal Society B. 276, 523–532.

Wiegand, T., Jeltsch, F., Hanski, I. and Grimm, V. (2003) Using pattern-oriented modeling for revealing hidden information:A key for reconciling ecological theory and application. Oikos 100, 209-222.

Willdenow, C.L. (1798) Grundriss der Kräuterkunde. Berlin. Haude & Spener.

Wissel, C. (1989) Theoretische Ökologie. Springer Verlag, Heidelberg.

Ziegenhagen, B., Gneuss, S., Rathmacher, G., Leyer, I., Bialozyt, R., Heinze, B., and Liepelt, S. (2008) A fast and simple genetic survey reveals the spread of poplar hybrids at a natural Elbe river site. Conservation Genetics 9, 373-379.

PUBLICATIONS AND MANUSCRIPTS

Paper 1

Distribution patterns of plants

explained by human movement behaviour

Ecological Modelling 220, 1339-1346

with J. Jetzkowitz, S. Brunzel, M.C. Wichmann and R. Bialozyt

Abstract

Distribution patterns of plants are affected by human activities such as creation, destruction or modification of habitats. However, another important question is to what extent humans shape plant distributions by acting as dispersal vectors. In order to answer this question we developed a simulation model for the spread of plant species between human settlements. This was done on the basis of extensive sociological and ecological data on a regional scale. With regard to the sociological data, human movement behaviour defined the amount of exchange between the settlements. Gardening types represented the potential habitat in our model. The ecological data was derived from a vegetation survey carried out in 2003, which was a repeat of a survey between 1974 and 1981 along the same transects. From these surveys, we studied the distributions of 13 species in 67 settlements. In our model, the earlier survey provided the data for the initial distribution. The simulated pattern was consequently compared with the distribution pattern in 2003. In the model, dispersal kernels based on patterns of human movement between settlements led to a better match with the distribution patterns than a null model simulating pure distance dependent dispersal for all species. This was statistically significant for seven of the thirteen species. A striking result was that alien species seem to benefit more from human dispersal than native species. We emphasize the importance of the sociological data on human movement behaviour in parameterising our regional scale model. This study provides quantitative evidence on the impact of human movement behaviour on the distribution of plant species in suburban areas.

1. Introduction

The ability to spread to their surroundings is an essential aspect of the ecology of all plant species, commonly referred to as dispersal (Bullock et al., 2002). Dispersal at different spatial scales enables plants to colonize new and distant habitats, to exchange genetic information among existing populations (maintaining genetic diversity), to support the survival of smaller populations (source-sink effects and metapopulations) or simply to maintain the spatial extent of the local population (Eriksson, 1996; Ouborg et al., 1999; Cain et al., 2000; Nathan, 2006). Natural dispersal processes can act at all spatial scales. Dispersal events over long distance are very infrequent and difficult to measure (see Bullock et al., 2006). Humans have a great potential as long distance dispersal vectors for plant and animal species, because of the high degree of transportation shown by humans (Hodkinson and Thompson, 1997).

Historical records document individual human mediated dispersal events over long distances. These include well known cases, such as the introduction of Himalayan Balsam (*Impatiens glandulifera*) and Japanese knotweed (*Fallopia japonica*) from Asia into Europe and North America (Beerling and Perrins, 1993; Hollingsworth and Bailey, 2000) as ornamental garden plants in the 19th century. These and similar are eye-catching cases of human mediated dispersal over large spatial scales that fundamentally change the pattern of species distribution. However, human mediated dispersal may also act more frequently at much smaller spatial scales with large impacts on species' distributions. These events act at the scale of kilometres and their impacts appear to be much more subtle and less evident than the examples of intercontinental dispersal given above. Nevertheless, the sum of numerous smaller-scale individual dispersal processes might be essential in shaping the distribution pattern of plants on intermediate scales.

While the impacts of human mediated dispersal at intermediate scales have been hypothesized by researchers for a long time (Ridley, 1930), this subject is intriguingly difficult to investigate as human movement behaviour is complex and many dispersal events will be accidental. Indeed, some studies give qualitative evidence with regard to everyday human activities that lead to seed dispersal by trains, cars or on human clothes (Hodkinson and Thompson, 1997; von der Lippe and Kowarik, 2007). In particular, the number of seeds and the distances they are carried by human activity have hardly ever been measured directly (but see Bullock et al., 2003) and only few studies provide insight into the actual biological process. Von der Lippe and Kowarik (2008) demonstrate that immense quantities of seeds are dispersed by cars, probably mostly due to airflow, while Wichmann et al. (2009) provide most detailed insight in seed dispersal on footwear. In particular, the latter authors measure how

many seeds exposed are then picked up on shoes, at what rate they fall off the shoe and how far they are carried (>5km). Dispersal by cars and by footwear, as well as the combination of the two may (among others) lead to the correlation patterns we find in our study.

Even beyond plants, very few studies have shown or have aimed to quantify the effects of human mediated dispersal An approach comparing patterns of distribution of an animal species with human movement behaviour (zebra mussel: Buchan and Padilla, 1999) or with human population density (horse chestnut leaf miner: Gilbert et al., 2004) has often found significant correlation and related models performed significantly better than simple diffusion models. This approach is used on plants in our study for the first time.

Plant species in settlements are affected by changes in these settlements. The decline of traditional farming and, in particular, the ongoing processes of urbanization, suburbanization and counterurbanization (Champion, 2001) have strong impacts on species habitats (e.g. Sharma and Amritphale, 2007). It must be emphasized that the transformation of rural settlements into urban landscapes around cities (Duany et al., 2000; Sieverts, 2003) should not only be considered as a conversion of rural to urban land cover (Bourne and Simons, 1982; U.S. Census Bureau, 2001) and therefore of the species habitats. Within this transformation the social organization and the lifestyles of the inhabitants of rural areas (Wirth, 1938; Popkin, 1999), including their patterns of mobility (Lefebvre, 1991; Urry, 2000), also change. Since the transformation of social systems can affect ecosystem processes like plant dispersal as direct or indirect drivers, many researchers explore the coupling of ecosystems to human action, especially in the field of land use and in urban ecology (Vitousek et al., 1997; Roy et al., 1999; Hill et al., 2002; Liu et al., 2003).

We scrutinise the effect of urbanization on plant distribution in settlements using a spatially explicit simulation model. This was tested by a comparison of the distribution pattern resulting from dispersal based on human movement behaviour versus the pattern generated by a commonly used dispersal kernel. Furthermore, we assessed the different importance of the urbanization factors, *i.e.* increased traffic and anthropogenic changes of habitat, for the distribution of native and alien plant species.

2. Methods

The main idea for modelling the dispersal process in this study was to analyze the influence of human beings on the distribution of plant species in settlements through two important processes. Firstly, they may create or destroy the habitat for the species by their gardening

activities, and secondly they disperse seeds and move adult plants according to their movement behaviour.

2.1. Dataset

The study area is located in Southern Hesse, Germany, in a region called "Wetterau". This region is located north of Frankfurt (Main). There are more than one hundred settlements within a total area of 1,101 km² and human population sizes range from only 150 up to 30,348 inhabitants. The distances between the settlements range from 1.2 to 36.1 km. The villages were characterised by rural life in former times but have increasingly become suburban residences. In particular, many people live in small suburban villages but work and spend their leisure time in bigger towns.

In 67 towns and villages a survey of plants along transects was carried out between 1974 and 1981 (Ludwig, unpublished data). There were three to six transects per settlement with an extent of 1,200 to 22,100 meters. This survey was repeated along the same transects in 2003. Brunzel et al. (2008) found an expanded range for many alien plants which, however, was not associated with a general decline of natives. During this survey the maximum suitable habitat for the establishment of ruderal plant species along transects within each settlement was estimated in m^2. Suitable habitats were classified by the presence of species of the phytosociological groups: *Sisymbrion officinalis*, *Eragrostion*, *Chenopodion rubri*, *Fumario-Euphorbion*, *Arction lappae* and *Polygonion avicularis* (after Ellenberg, 1996; Oberdorfer, 1983; see also Pyšek et al., 2004). This variable, therefore, quantifies the area that is not covered by asphalt, lawn or woodland.

For the modelling approach, 13 plant species were selected. The selection was based on the criterion that the species had to be moderately frequent but not common in the former survey and in 2003. This was an important prerequisite for simulating dispersal, because these species showed a clear change in the distribution between the former survey and 2003.

At the same time as the plant survey in 2003, a sociological study was carried out on the life- and gardening styles within these 67 settlements. Questionnaires were developed to collect data on gardening styles and practices as well as on household mobility. All together 1,359 households were polled, selected at random within the 67 settlements. The classification of the gardening styles into nine gardening types were carried out by means of lifestyle analysis (Jetzkowitz et al., 2007) which describes how people in the region use their gardens and the space around their dwelling (Table 1). The proportions of the nine gardening types per settlement were used to populate a habitat matrix with the settlements and gardening types in

the rows and columns respectively. Additionally, two combinations of gardening types (1 – 5 and 6 – 9) and two fixed values for all settlements (50%, 100%) were added to the habitat matrix to test if these artificial arrangements and combinations represent a species' habitat better. This resulted in thirteen different habitats to be used in our modelling approach (Table 1).

The potential amount of exchange of seeds and adult plants between settlements was defined by characterizing the household mobility of a settlement. This mobility was calculated according to the "Spatial-Temporal Action and Movement Pattern" (STAMP) concept (Jetzkowitz et al., 2007; see also supplementary). The concept describes the total distances individuals travel for work, shopping and during leisure time. We transformed the movement data into a matrix of connectivity between the 67 settlements (see supplementary). The entries in this symmetric transport matrix M_{trans} are the number of people moving between these settlements. It has to be mentioned that seeds and especially adult plants are not only carried by humans by chance but also intentionally for ornamental purposes. This was not explicitly included in the transport matrix because human behaviour is independent of plant distribution and characteristics. Though, ornamental value may help to explain the observed distribution pattern.

Table 1 – The used gardening types and their characteristics

Gardening types	**Characteristics**
No garden	Household has no garden
Recreational garden	Place to relax with equipment for recreation
Lawn garden	No planted beds, only grass
Traditional fruit and vegetable garden	Traditional cultivation with beds for vegetables and fruits
Flower garden	Garden with flowers and ornamental plants
Event garden	For representation and other events
Sealed garden	Most area sealed with paving
Child-friendly garden	Family garden with equipment for children
Traditional All-in-One garden	Traditional cultivation, flower beds and decoration
Combination 1	The sum of the proportions of garden 1 to 5
Combination 2	The sum of the proportions of garden 6 to 9
Null-Hypotheses 1	50% suitable habitat in the settlement
Null-Hypotheses 2	100% suitable habitat in the settlement

2.2. Model

We used a spatially explicit simulation model whereby the studied 67 towns were arranged according to their geographical coordinates. By doing so, a landscape for the dispersal process emerged. A null model was created by simulating plant dispersal using a Gaussian function on the geographical distances between the towns. This was represented in a matrix (M_0) to simulate pure distance-dependent dispersal. Sensitivity analysis revealed no improvement of the results using standard deviations of the Gaussian function between 0.5 and 10 km. Therefore, the standard deviation of the Gaussian function was set to 5 km to allow stepwise colonization of every location from just a single source. On average, each town had 5.7 neighbours within a range of 5 km. Although this is a fat tailed distribution, long distance dispersal events over more than 20 km are still rare. The habitat matrix and the two dispersal matrices M_{trans} and M_0 were used to parameterize the habitat availability and the dispersal process respectively.

2.2.1. Assumptions

The aim of modelling is always to simplify by focusing on the essential parts and processes. For this reason, we make the following simplifying assumptions in the model:

[1] The gardening styles and the household mobility do not change during the simulation time steps.

[2] There is a carrying capacity for every town based on the estimation of maximum suitable habitat for ruderal plant species during the survey along transects in 2003 and is the same for all species.

[3] We assume gardening types to drive available habitat. The amount of habitat for a single species in a town is restricted to the proportion of a gardening type within this town multiplied with the carrying capacity of the town.

[4] The species disperse only once per time step (year).

[5] A species-specific mortality rate (r_{spec}) reduces the number of populations of this species per town once per year (e.g. winter or human mediated mortality). This was necessary because we did not simulate population dynamics at the local scale.

[6] A species-specific dispersal coefficient (c_{spec}) represents the capability for dispersal of the given species. In nature, this could be based on morphological traits for seed dispersal such as hooks, but also having attractive flowers so that these plants are transported and planted into new gardens.

2.2.2. Dispersal process

In our model dispersal is simulated sperately for each species and habitat type. The dispersal between settlements is dependent on the amount of exchange between the settlements (M_{trans} or M_0) and the species-specific dispersal coefficient (c_{spec}, derived by a genetic algorithm as described below). Additionally, the model considers the amount of habitat in the destination settlement j , i.e. $habitat_j$, because successful dispersal on the population scale includes establishment of new populations. The number of new populations P^{new} at settlement j is

$$P_j^{new} = \sum_{i=1}^{67} \left(P_i \ * (M_{trans} \ or \ M_0)_{ij} * c_{spec} * habitat_j \right) \quad (1)$$

with i giving the settlement of origin.

Because the model is deterministic, and therefore P_j is a rational number, the population size P_i in the source settlement for dispersal in the next time step was rounded down to the next integer value. Herewith only complete populations are used for the dispersal process. For the case $i=j$ (inner-settlement colonisation) only P_i and the amount of habitat is considered for the establishment of a new population.

2.2.3. Simulation

The distribution of a plant species in the years 1974 to 1981 (Ludwig, unpublished data) was used as its initial distribution. The simulations ran for 20 years (time steps). At each time step every population in each town could disperse to all other towns according to the dispersal process (equation 1). At the end of each time step the number of populations was reduced in each town dependent on the species-specific mortality rate (r_{spec}, derived by a genetic algorithm as described below). After 20 time steps the resulting species distribution was compared with the natural distribution in 2003. The deviation of the two patterns was calculated as a measure for the quality of the modelled dispersal and establishment process. This measure takes into account the distance between settlements populated by the species in nature and in simulations. If a settlement i was populated at the simulation end although it was not populated in nature, the linear distance to the next naturally populated settlement yielded the deviation d(nat, sim). These deviations for all populated settlements were summed to a total deviation. This procedure was then repeated for the settlements j that were naturally populated but not at simulation end d(sim, nat). The mean of the two deviations defined the total deviation (d_{total}) for one simulation run (see Fig. 1).

$$d_{total} = \left(\frac{\sum_{i \in NAT} \min_{k \in SIM} d(nat_i, sim_k) + \sum_{j \in SIM} \min_{k \in NAT} d(sim_j, nat_k)}{2} \right) \quad (2)$$

d(nat, sim) is the geographic distance, SIM is the set of all populated settlements after the simulation and NAT is the set of all naturally populated settlements.

In case of a perfect fit d_{total} would be zero. Hence, d_{total} provides a measure of mismatch between the modeled and natural patterns. The simulations were repeated for all 13 species in combination with the 13 habitats based on the habitat matrix.

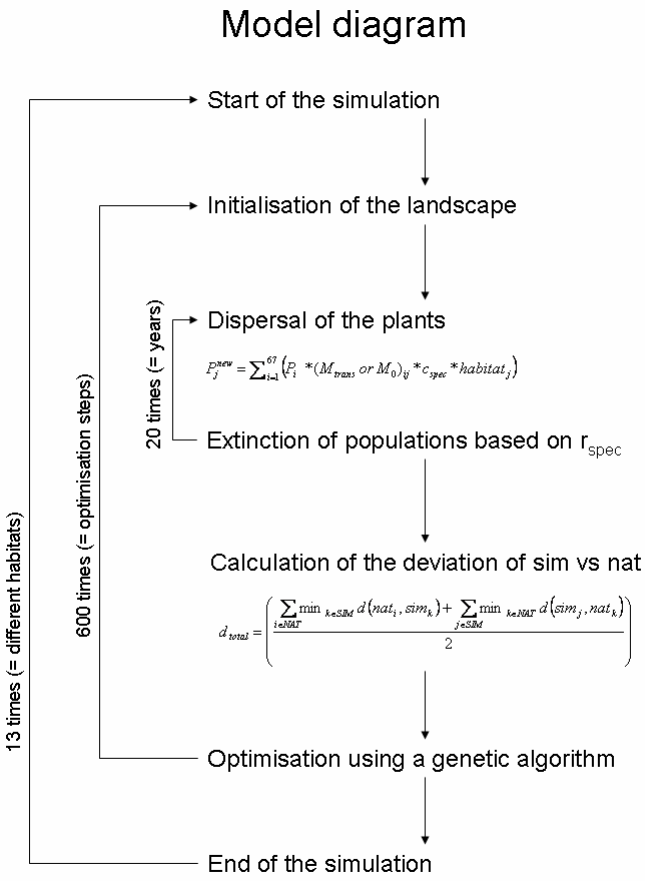


Fig. 1 – The conceptual diagram of the simulation model showing the different steps from the initilisation of the paramters to the calculation of the deviation d_{total} as a measure of mismatch between simulated and natural plant distribution. This was done for all analysed species. Abbreviations are explained in the methods section 2.2.2 and 2.2.3.

2.2.4. Genetic Algorithm

Empirical studies have not yet provided on the potential of human mediated dispersal and therefore the species-specific dispersal coefficient (c_{spec}), as well as the species-specific mortality rate (r_{spec}) were estimated. These two parameters were then optimized for a minimum total deviation (d_{total}) for both matrices in order to be able to make a comparison of the pattern generated using either matrix (M_{trans}, M_0). This was done by using a Genetic Algorithm (GA).

GA's became widely recognized as an optimization method to mimic the selection process during evolution (Holland, 1975). The unit of selection is the so-called chromosome on which the values for the simulations (in our case c_{spec} and r_{spec}) are located. To generate new chromosomes (combinations of c_{spec} and r_{spec}, respectively) crossovers and mutations are allowed. In our case the fitness of the chromosome is the d_{total} of the simulation using these values.

The GA was initialized with 24 combinations of c_{spec} and r_{spec} covering the range of possible values for dispersal and mortality rate. With each combination a simulation run was performed to obtain the first total deviations d_{total} for the simulated species. The values of c_{spec} and r_{spec} of the twelve combinations resulting in the smallest and therefore the minimal deviation d_{total} were mutated to produce twelve new combinations that replaced the twelve combinations with the highest d_{total}. The GA ran for 600 time loops with decreasing mutation rates. This was done for all combinations of species, habitat and dispersal matrices separately.

2.3. Statistical Analysis

All statistical analyses were carried out with SPSS 11.0 and R 2.0.1 (SPSS Inc., 2001; R-Development Core Team, 2004). The effects of the three categorical variables (species, habitat and dispersal matrix [M_{trans}, M_0]) on the simulation result (dependent variable d_{total}) were analyzed by applying a generalized linear model. Afterwards the analysis was done for all 13 species separately. As two of the nine gardening types produced in some species an exceedingly high deviation, these two types (type 3 and 4) were not used in the further analysis.

3. Results

The modelling results showed for all analyzed species a closer or equal match with regard to the natural distribution over all habitats using the sociologically-inspired human transport matrix M_{trans} than using the best fitting null model M_0 (Fig. 2).

Applying the generalized linear model to all species together, d_{total} was significantly affected by the species ($F_{12,262} = 44.02$, $p < 0.001$) and the dispersal matrix ($F_{1,262} = 57.39$, $p < 0.001$). The habitat had no significant effect on d_{total}.

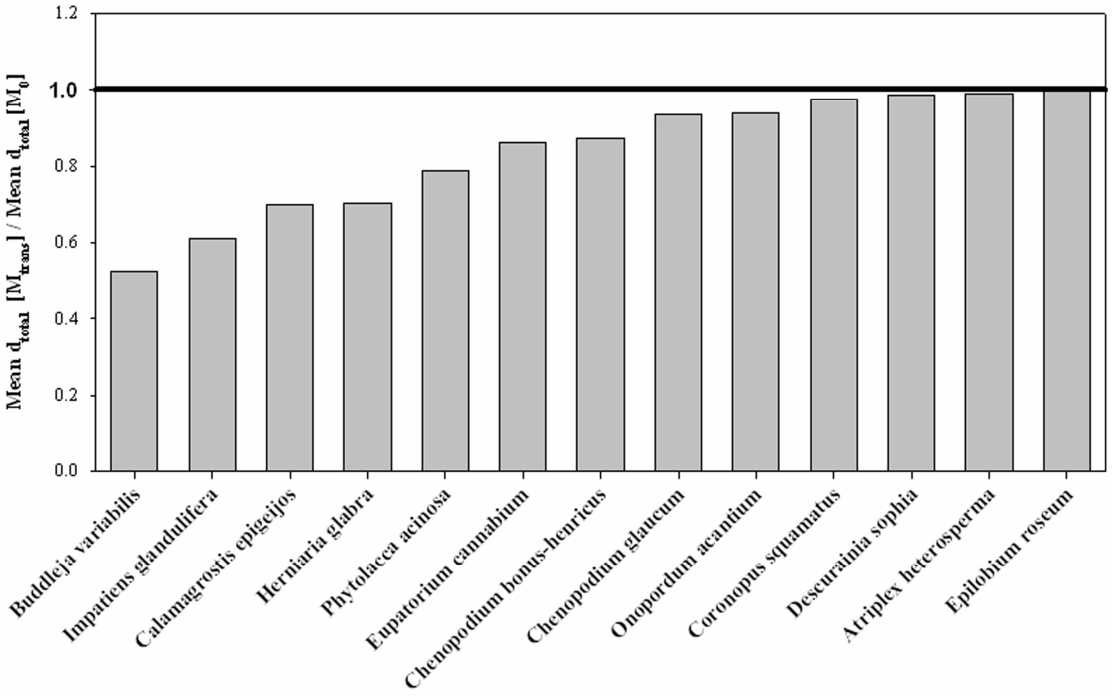


Fig. 2 – Mean of deviation d_{total} for each species across all analyzed habitats (gardening types) using the sociologically-inspired transport matrix M_{trans} divided by the mean of deviation d_{total} for each species across all analyzed habitats using the matrix M_0 of the null model. As d_{total} is a measure of mismatch between simulated and natural plant distribution, values below 1 indicate that the deviation of M_{trans} is lower than the one of M_0. Note, that in no case did M_0 give a better fit.

Analysis of the 13 species separately, showed significant effects of the habitat for nine and significant effects of the dispersal matrix for seven (Table 2). In only a few cases the additional combination of several gardening types (1-5 or 6-9) or one of the fixed habitats (50% or 100% of suitable habitat) produced a better result than any of the single habitats alone. In most cases the additional types produced values for d_{total} somewhere in the range of the results of the single habitats (Fig. 3). It was also not possible to group the species according to their best-performing habitat into functional groups.

Comparing the effects of habitat and dispersal matrix among species, some species were more affected by the habitat (Fig. 3a), whereas others showed a clear sensitivity to the dispersal matrix (Fig. 3b). In the first group we found mostly native species whereas alien species dominated the second group (see also Table 2).

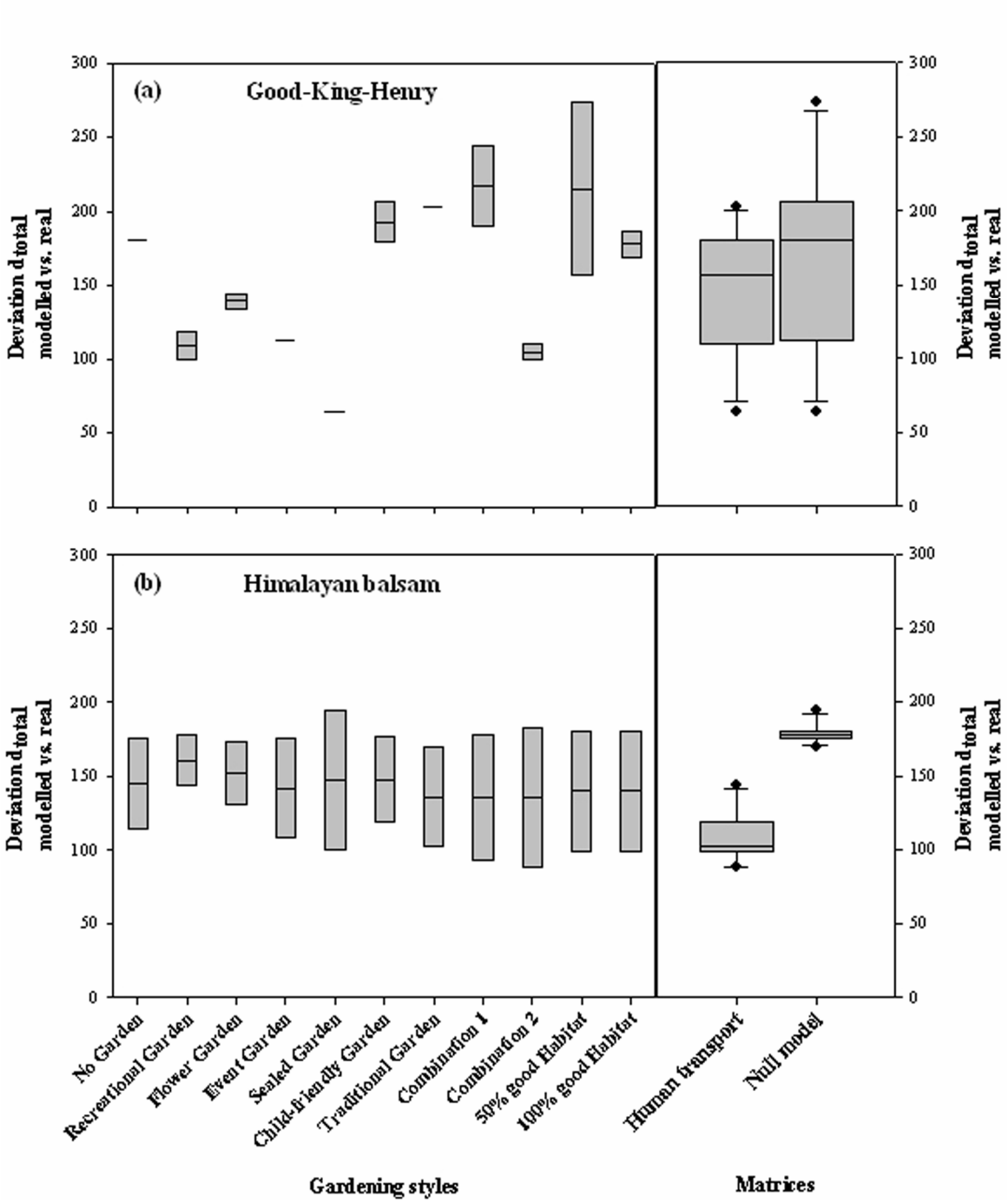


Fig. 3 – Comparison of the effect of habitat (gardening type) and dispersal matrices on (a) Good-King-Henry (*Chenopodium bonus-henricus)* and (b) Himalayan balsam *(Impatiens glandulifera)* by the deviation d_{total}. The native *Ch. bonus-henricus* is most affected by the habitat whereas the alien species *I. glandulifera* is most affected by the dispersal matrices. Note, d_{total} is a measure of mismatch between simulated and natural plant distribution. Low values in d_{total} indicate model scenarios that are more likely to mirror natural processes.

Table 2 – Results of analysis of variance for all species separately

Species	Factor	df	MS	F	
Atriplex	Habitat	10	5150.866	39.899	***
heterosperma	Dispersal matrix	1	4.855	0.038	ns
(A)	Error	10	129.096		
Buddleja	Habitat	10	1921.899	1.182	ns
variabilis	Dispersal matrix	1	105860.755	65.091	***
(A)	Error	10	1626.345		
Calamagrostis	Habitat	10	1117.570	3.522	*
epigeijos	Dispersal matrix	1	22831.829	71.955	***
	Error	10	317.305		
Chenopodium	Habitat	10	5464.857	8.437	**
bonus-henricus	Dispersal matrix	1	2488.698	3.842	ns
	Error	10	647.747		
Chenopodium	Habitat	10	2462.977	46.292	***
glaucum	Dispersal matrix	1	253.667	4.768	ns
	Error	10	53.205		
Coronopus	Habitat	10	3584.277	14.974	***
squamatus	Dispersal matrix	1	88.997	0.372	ns
	Error	10	239.370		
Descurainia	Habitat	10	674.822	11.623	***
sophia	Dispersal matrix	1	3.741	0.064	ns
	Error	10	58.060		
Epilobium	Habitat	10	1809.724	32.224	***
roseum	Dispersal matrix	1	16.500	0.294	ns
	Error	10	56.161		
Eupatorium	Habitat	10	1627.626	1.325	ns
cannabium	Dispersal matrix	1	6426.697	5.233	*
	Error	10	1228.155		
Herniaria	Habitat	10	1531.532	4.089	*
glabra	Dispersal matrix	1	27373.687	73.080	***
	Error	10	374.572		
Impatiens	Habitat	10	143.234	0.645	ns
glandulifera	Dispersal matrix	1	25791.069	116.191	***
(A)	Error	10	221.972		
Onopordum	Habitat	10	498.810	2.917	ns
acanthium	Dispersal matrix	1	876.832	5.128	*
	Error	10	170.989		
Phytolacca	Habitat	10	2400.161	6.130	**
acinosa	Dispersal matrix	1	13362.234	34.127	***
(A)	Error	10	391.547		

Alien species are marked (A). Levels of significance are: *** $p<0.001$; ** $p<0.01$; * $p<0.05$.

4. Discussion

With our spatially explicit simulation model we analysed the effect of urbanization on plant distributions in settlements. Human mediated dispersal between and the amount of habitat within settlements was modelled using sociological and ecological data. By comparison of the simulated patterns based on human behaviour or a pure distance dependent kernel with the natural distribution in 2003 we could rank the outcome of our simulations.

The general conclusion of our modelling study is that plant distribution is strongly impacted by human behaviour. This is in line with earlier findings (e.g. Hope et al., 2003; Antrop, 2004) which suggest that urbanization rewrites the pattern of plant distribution. With increasing human mobility and urbanization humans have become increasingly responsible for plant distributions seen today (Antrop, 2004). Several authors recognized that diversity is higher within cities than in the adjacent countryside (Pyšek, 1993; Kuhn et al., 2004). The explanation for this is (i) the higher heterogeneity of the habitat within cities and (ii) the high probability for new species arriving there through transportation processes (Sukopp et al., 1979; Kowarik, 1990; von der Lippe and Kowarik, 2007). These two factors are the two constraints we have in mind when devising our simulation study. Habitat heterogeneity between cities is represented by the gardening styles and transportation is represented by the human mobility, realized using the sociological data and the STAMP concept (Jetzkowitz et al., 2007).

Our results demonstrate high impacts of human mediated dispersal on the distribution of plants (e.g. Fig. 4). This is most effectively demonstrated by the ratio of the deviation d_{total} from the human transport matrix M_{trans} over the matrix of the null model M_0. In all cases this ratio is lower than one and therefore human transport explains plant distribution better than our best fitted null model (Fig. 2).

In principle it is known that humans influence the dispersal of plants. Seeds might be attached to cars, human clothes or to the hair of their dogs (Salisbury, 1942; Clifford, 1956; Lonsdale and Lane, 1994; Graae, 2002; Zwaenepoel et al., 2006; Wichmann et al., 2009). Furthermore attractive plants or their seeds are carried intentionally for garden planting (e.g. Beerling and Perrins, 1993). Here we have been able to show the consequences of the sum of such dispersal processes on a regional scale. This was done through a simulation model using real data on human movement behaviour. Although the human transport matrix was invented to investigate household mobility in urbanized regions, it explained the change in plant species distribution between 1981 and 2003 as well as or better than our null model, the best fitted distance dependent dispersal kernel. All analyzed species showed the same trend. This shows

the considerable impact of human movement behaviour on the dispersal process in urbanized regions. In particular, humans can transport seeds from one good habitat to another one irrespective of the type of intervening landscape and also over long distances. It has been shown that traffic acts as a long distance vector for plant dispersal (von der Lippe and Kowarik, 2007). Such long distance events shape the species distribution. Gaps within the distribution may be explained by the targeted dispersal process leaping such gaps. An alternative explanation for the gaps may be fragmentation of habitats where not all connections are realized. Wind or animal dispersal has to take many steps (i.e. using different habitat patches as 'stepping stones') to cover large distances whereas human mediated dispersal may cover the same distance in a single step. This may also account for the deviation of the null model from the natural distribution (Fig. 4). Furthermore, the species studied mostly showed an increase in their range. Species which reduced in range during the simulation showed no difference between the two dispersal matrices. We would expect this if the final range of the species is a subset of the former range. Therefore, dispersal is not such an important process for declining species in our simulations.

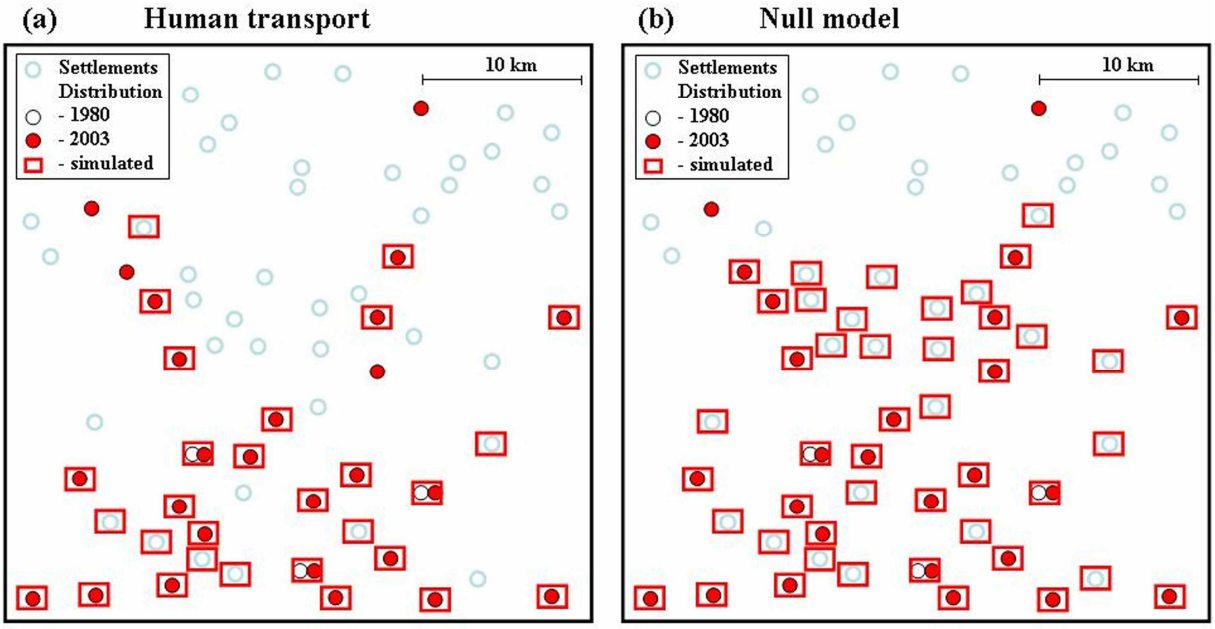


Fig. 4 – Visiualising one example of comparing the best simulation results using (a) the sociologically-inspired human transport matrix and (b) the null model for one exemplary gardening style (Recreational garden) and Himalayan balsam (*Impatiens glandulifera*). Note, our analysis indicates lower deviation of the simulated from the natural distribution 2003 when using the human transport matrix (a) compared to the null model (b).

In many investigations concerning plant species in settlements special attention is given to the differences between alien and native species in terms of establishment and dispersal (Lonsdale, 1999; Maskell et al., 2006; von der Lippe and Kowarik, 2008). In our investigation we used both types of species in the simulations to address this point. Interestingly, native plants tended to show a stronger response to the habitat than to the dispersal scenario. This is exemplified in Fig. 3a for Good-King-Henry (*Chenopodium bonus-henricus*) and agrees with the hypothesis that species distribution is habitat-dependent (Funk and Vitousek, 2007). Some species, such as Himalayan balsam (*Impatiens glandulifera*) tended to show stronger response to the dispersal scenario than to the habitat (Fig. 3b). We found such behaviour mostly in alien species (75% of the alien species vs. 44% of the native species). This suggests that human movement behaviour may have less effect on natives but that aliens are more adapted to humans as transport vectors (see also von der Lippe and Kowarik, 2007; Brunzel et al., 2008).

With our approach we considered the origin and the destination of a seed or adult plant, but not its individual pathway through the landscape. Humans may disperse seeds not only on their way from home to work and vice versa but also to the places where they have a break during their journey. Furthermore, they may interact with others and accidentally exchange seeds for example at a bus stop. Therefore an agent-based model could reveal further insights into the effects of human movement patterns, especially with regard to new established public transport routes between settlements.

Although our sociologically-inspired human transport matrix does contribute significantly to the explanation of plant distribution (Fig. 2, 3b), one might criticize it in that it is symmetric. Dispersal may be highly directed, e.g. from the main town to suburban areas (von der Lippe and Kowarik, 2008) where the agricultural dominated society and habitat structure is changing into an urban environment. In fact, direction could be a very interesting parameter for a more detailed study, but our sociological dataset does not include this information.

Because the dataset was not designed for a modelling approach, we had to compensate the missing data by creation of a robust model. The major aim was the use of human movement behaviour and garden activities for plant species dispersal in an unique and effective way. Therefore we had to make some assumptions.

Assumption [1] state that there is no change in gardening style and household mobility between 1980 and 2003. Of course the human behaviour of 2003 was different to the one in the 1980s. Nevertheless we used the human behaviour of 2003 for all years in our simulation. It is assumed that during the studied time period the changes have been one-directional

towards the situation that we see today (Jetzkowitz et al., 2007) and the process is still ongoing. Because the dispersal process in our simulation depends not only on the human movement and the gardening type in the target settlement but also on the optimized dispersal coefficient, the increase of connections between settlements and special gardening types is compensated by this factor. Therefore our simulation results and the strong impact of human behaviour onto the plant distribution would have been even more distinct if we had data for the whole period.

Assumption [5] and [6] are dealing with the death rate and the dispersal coefficient. Both values are unknown and are optimized during the simulation to compare the best fitted distance dependent null model with the best one using the sociological inspired human transport matrix. Because we want to show the overall importance of human movement behaviour on the distribution of plant species, our procedure seems to be an appropriate solution. The exact values for r_{spec} and c_{spec} could be hardly compared with real data, because we used these values for overall coefficients including many different processes and traits.

5. Conclusion

Our study highlights the strong impact of human movement behaviour on the distribution of plants in settlements. The study shows the relevance of introducing sociological data into ecological research in human dominated landscapes. As we could show, information on human movement behaviour is necessary to forecast the distribution of plant species in settlements, especially for alien species.

Acknowledgement

We thank J.M. Bullock for comments on the manuscript; J. Schneider and S. Fischer for the work on the sociological and biological data; and T. Fiedler for the transport matrix. This work was funded by the Volkswagen Foundation (II/77 443-1).

References

Antrop, M., 2004. Landscape change and the urbanization process in Europe. Landscape and Urban Planning 67, 9-26.

Beerling, D.J., Perrins, J., 1993. *Impatiens glandulifera* Royle (*Impatiens roylei* Walp.). Journal of Ecology 81, 367-382.

Bourne, L.S., Simmons, J.W., 1982. Defining the area of interest: definition of the city, metropolitan areas and extended urban regions, in: Bourne, L.S. (Ed.), Internal structure of the city. Oxford University Press, New York, 57-72.

Brunzel, S., Fischer, S.F., Schneider, J., Jetzkowitz, J., Brandl, R., 2008. Neo- and archaeophytes respond more strongly than natives to socio-economic mobility- and disturbance patterns along an urban-rural gradient. Journal of Biogeography, doi:10.1111/j.1365-2699.2008.02044.x.

Buchan, L.A. J., Padilla, D.K., 1999. Estimating the probability of long-distance overland dispersal of invading aquatic species. Ecological Applications 9, 254-265.

Bullock, J.M., R.E. Kenward, Hails, R.S., 2002. Dispersal Ecology. Blackwell Science, Oxford.

Bullock, J.M., Moy, I.L., Coulson, S.J., Clarke, R.T., 2003. Habitat-specific dispersal: environmental effects on the mechanisms and patterns of seed movement in a grassland herb Rhinanthus minor. Ecography 26, 692-704.

Bullock, J.M., Shea, K., Skarpaas, O., 2006. Measuring plant dispersal: an introduction to field methods and experimental design. Plant Ecology 186, 217-234.

Cain, M.L., Milligan, B.G., Strand, A.E., 2000. Long-distance seed dispersal in plant populations. American Journal of Botany 87, 1217-1227.

Champion, T., 2001. Urbanization, Suburbanization, Counterurbanization and Reurbanization. In: Paddison, R. (ed.) Handbook of Urban Studies, Sage, London, 143-161.

Clifford, H.T., 1956. Seed dispersal on footwear. Proceedings of the Botanical Society of the British Isles 2, 129-131.

Duany, A., Plater-Zyberk, E., Speck, J., 2000. Suburban Nation. The Rise of Sprawl and the Decline of the American Dream. North Point Press, New York.

Eriksson, O., 1996. Regional dynamics of plants: A review of evidence for remnant, source-sink and metapopulations. Oikos 77, 248-258.

Funk, J.L., Vitousek, P.M., 2007. Resource-use efficiency and plant invasion in low-resource systems. Nature 446, 1079-1081.

Gilbert, M., Gregoire, J.C., Freise, J.F., Heitland, W., 2004. Long-distance dispersal and human population density allow the prediction of invasive patterns in the horse chestnut leafminer Cameraria ohridella. Journal of Animal Ecology 73, 459-468.

Graae, B.J., 2002. The role of epizoochorous seed dispersal of forest plant species in a fragmented landscape. Seed Science Research 12, 113-120.

Hill, M. O., Roy, D. B., Thompson, K., 2002. Hemeroby, Urbanity and Ruderality: Bioindicators of Disturbance and Human Impact. Journal of Applied Ecology 39, 708-720.

Hodkinson, D.J., Thompson, K., 1997. Plant dispersal: the role of man. Journal of Applied Ecology 34, 1484-1496.

Holland, J., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press, Michigan.

Hollingsworth, M.L., Bailey, J.P., 2000. Evidence for massive clonal growth in the invasive weed Fallopia japonica (Japanese Knotweed). Botanical Journal of the Linnean Society 133, 463-472.

Hope, D., Gries, C., Zhu, W., Fagan, W.F., Redman, C.L., Grimm, N.B., Nelson, A.L., Martin, C., Kinzig, A., 2003. Socioeconomics drive urban plant diversity. Proceedings of the National Academy of Sciences USA 100, 8788-8792.

Jetzkowitz, J., Schneider, J., Brunzel, S., 2007. Suburbanisation, Mobility and the "Good life in the country". A lifestyle Approach to the Sociology of Urban Sprawl in Germany. Sociologia Ruralis 47, 1-24.

Kowarik, I., 1990. Some responses of flora and vegetation to urbanization in Central Europe. In: Sukopp, H., Hejny, S., Kowarik, I. (eds), Urban ecology, SPB Academic Publishing, The Hague, pp. 45-74.

Kuhn, I., Brandl, R., Klotz, S., 2004. The flora of German cities is naturally species rich. Evolutionary Ecology Research 6, 749-764.

Lefèbvre, H., 1991. The Production of Space. Blackwell, New York.

Liu, J., Daily, G. C., Ehrlich, P. R., Luck, G. W., 2003. Effects of Household Dynamics on Resource Consumption and Biodiversity. Nature, 421, 530-533.

Lonsdale, W.M., Lane, A.M., 1994. Tourist vehicles as vectors of weed seeds in Kakadu National Park, Northern Australia. Biological Conservation 69, 277-283.

Lonsdale, W.M., 1999. Global patterns of plant invasions and the concept of invasibility. Ecology 80, 1522-1536.

Maskell, L.C., Firbank, L.G., Thompson, K., Bullock, J.M., Smart, S.M., 2006. Interactions

between non-native plant species and the floristic composition of common habitats. Journal of Ecology 94, 1052-1060.

Nathan, R., 2006. Long-distance dispersal of plants. Science 313, 786-788.

Ouborg, N.J., Piquot, Y., van Groenendael, J.M., 1999. Population genetics, molecular markers and the study of dispersal in plants. Journal of Ecology 87, 551-568.

Popkin, B.M., 1999. Urbanization, lifestyle changes and the nutrition transition. World Development 27, 1905-1999.

Pyšek, P., 1993. Factors affecting the diversity of flora and vegetation in central-european settlements. Vegetatio 106, 89-100.

The R-Development Core Team, 2004. R: A Language and Environment for Statistical Computing. Available online at: http://www.R-project.org.

Ridley, H.N., 1930. The dispersal of plants throughout the world. L. Reeve & Co, Ashford, Kent.

Salisbury, E.J., 1942. The weed problem. Nature 149, 594-596.

Sharma, S., Amritphale, D., 2007. Effect of urbanization on the population density of aak weevil. Current Science 93, 1130-1134.

Sieverts, T., 2003. Cities without Cities: Between Place and World, Space and Time, Town and Country, Routledge, London.

SPSS Inc. 2001. SPSS 11.0 for Windows. Chicago, IL.

Sukopp, H., Blume, H.-P., Kunick, W., 1979. The soil, flora and vegetation of Berlins waste lands. In: Laurie, I.E. (ed), Nature in cities, John Wiley & Sons, Chichester, pp. 115-31.

Urry, J., 2000. Mobile sociology. British Journal of Sociology 51, 185-203.

US Census Bureau, 2001. Statistical Abstract of the United States. Government Printing Office, Washington DC.

Vitousek, P. M., Mooney, H. A., Lubchenco, J., Melillo, J. M., 1997. Human Domination of Earth`s Ecosystems. Science, 277, 494-499.

von der Lippe, M., Kowarik, I., 2007. Long-Distance Dispersal of Plants by Vehicles as a river of Plant Invasions. Conservation Biology 4, 986-996.

von der Lippe, M., Kowarik, I., 2008. Do cities export biodiversity? Traffic as dispersal vector across urban–rural gradients. Diversity and Distributions 14, 18-25.

Wichmann, M.C., Alexander, M.J., Soons, M.B., Galsworthy, S., Dunne, L., Gould, R.,

Fairfax, C., Niggemann, M., Hails, R.S., Bullock, J.M., 2009. Human-mediated dispersal of seeds over long distances. Proceedings of the Royal Society B. 276, 523-532.

Wirth, L., 1938. Urbanism as a Way of life, Am. J. Sociol. 40, 1-24.

Zwaenepoel, A., Roovers, P., Hermy, M., 2006. Motor vehicles as vectors of plant species from road verges in a suburban environment. Basic and Applied Ecology 7, 83-93.

Supplementary

Obtaining the human movement rates

The calculation of human movement rates are based on the concept of "Spatial-Temporal Action and Movement Patterns" (STAMPs) developed by Jetzkowitz et al. (2007). In answer to the question of what a person does where, the STAMP concept distinguishes only between work (including house work, training, military service, etc.) and leisure. Data ascertained in the survey were transformed into 17 variables, which record six differing mobility aspects (as dimensions of the results of factor analyses or as z-standardised individual variables):
(1) everyday activity locations (leisure and consumption radii); (2) unusual leisure activity locations (holiday radii); (3) residential locations of the family; (4) use of a car; (5) variables characterizing the residential situation; (6) attachment to place of residents.
By means of an explorative cluster analysis nine STAMPs are identified:

No	STAMP	Percentage	Main Characteristics
1	Bound to urban life	(8%)	Leisure mobility to Frankfurt or another city in the region Employment in Frankfurt
2	Commuting for work	(8%)	Employment in Frankfurt
3	Home-bound	(15%)	Working at home No or less vacations Garden as preferred leisure location
4	Mobile without relation to the locality	(10%)	Working in the region Weak attachment to the locality
5	Employment in the locality	(12%)	Employment on the spot Not working at home or in the region
6	Mobile for retreat	(18%)	Working at home Many vacations Leisure in the garden
7	Employment distantly	(9%)	Employment outside the region Above-average distance to work place
8	Bound to the locality and the region	(13%)	Employment in the region Family in the vicinity of the place of residence Strong attachment to the locality
9	Immobile	(8%)	Staying at home

Movement patterns were calculated on the basis of these classification taking into account two destined locations, the place of work and the place of leisure activities. The movement patterns describe both the frequency of how many people representing a certain STAMP really go to the specified locations and the maximum distance the people tolerate to reach that location. The percentages of people travelling to work can be extracted from the empirical database. The percentages of people spending their leisure in one of the destined settlements had to be recalculated (scaled and normalized) from factor values obtained when the answers in the questionnaire were clustered. The means of both percentages result in the frequencies in the movement patterns. By staying at home or in the hometown the radius is zero. By going to

a regional city the radius of this movement was identified as 11.9 km. This radius was the result of the analyses of the human movement behaviour, with the highest percentage of people going to locations within the region.

Finally, the supply infrastructure has to be taken into account to calculate the human movement rates. Locations within the region were classified as with full, medium or no supply infrastructure. Without such an infrastructure, towns or villages could not act as destination for human movement. Locations with medium supply infrastructure were visited by people living in locations within the valid radius of 11.9 km. Destinations with a full supply infrastructure could be travelled to if they were located within a radius 1.5fold the valid radius: 17.8 km. This classification expresses the attractiveness of a location. In order to choose from a wider variety of products in supermarkets or in warehouses or to have more choices of free-time activities, people are assumed to accept greater distances to travel.

The STAMPs and their distribution, the movement patterns, the supply infrastructure and the number of inhabitants were then used to calculate the human movement rates. They state for every two locations the number of humans commuting between them. The function *f* takes one location *i* and a STAMP as input and returns the movement rates only for that input location in a matrix the size of the full movement rates matrix, where unused parts are set to zero. The full matrix can be obtained by summing-up the single matrices over all locations and all STAMPs: $W = \sum_{i \in Towns} \sum_{s=1}^{9} f(i,s)$, where W is the full movement-rates matrix and s enumerates the nine STAMPs. By counting settlements with medium supply infrastructure within the valid radius of 11.9 km and locations with full infrastructure within the 1.5fold valid radius the number of locations, that could be reached, is obtained. As the final step the function *f* calculates the movement rates based on the input locations. Those humans willing to travel are equally distributed on all locations of a "place of destination" class.

An example might illustrate this more clearly: The village "Windecken" comprises 1000 inhabitants. One tenth of the population belongs to STAMP 4. Within STAMP 4 50% go to "locations within the region". Let 10 locations of the region are counted within the different radii. The movement rate to each location due to STAMP 4 is $\frac{(1000 \cdot 0.1 \cdot 0.5)}{10} = 5$.

By using the above described calculation method a matrix W of stored human movement rates is obtained (see supporting Excel file). A connection between two towns is defined "established" if the corresponding matrix entry has a value greater than zero and therefore people travel between the two towns. Thus, a connection between Altenstadt and Bisses is established, but not between Assenheim and Borsdorf.

Paper 2

Marked point pattern analysis on genetic paternity data for estimation and uncertainty assessment of pollen dispersal kernels

Submitted

with T. Wiegand, J.J. Robledo-Arnuncio and R. Bialozyt

Summary

1. The estimation of a suitable pollen dispersal kernel is essential to forecast the consequences of gene flow between and within populations.

2. We analyzed the spatial structure of pollen dispersal and estimated pollen dispersal kernels for a population of *Populus nigra* L. located in Central Germany based on results of genetic paternity analyses. Marked point pattern analysis allowed us to account for the spatial structure of males and females and to assess the uncertainty in the data at different distance intervals.

3. Our analysis showed a significant departure from the null model, obtained by random shuffling of the father trees, up to a distance of approximately 300m. Simulation envelopes of the null model were uncomfortably wide at larger distances indicating large uncertainty on the detailed shape of the kernel's tail.

4. Model selection revealed that two-component pollen dispersal kernels comprising an exponential power function and a truncated uniform function were the most parsimonious models to fit the data. The fitted kernel provided comparable results with published spatial statistic models. Kernel estimates based on the direct fit of the observed mating distance distribution differed strongly from consistent results obtained by methods that account for the spatial structure in the population.

5. Synthesis: Spatial structure of pollen sources and sinks must be accounted for when deriving dispersal kernels. The proposed marked point pattern analysis accounts for this structure and allows for an assessment of the uncertainty in kernel estimates. Our results indicate that statistical model fitting may substantially underestimate the uncertainty in kernel estimation, especially at larger distances.

Introduction

Gene flow among and within plant populations concerns evolutionary ecologists, conservationists and ecosystem managers. Spatial patterns of pollen and seed dispersal determine landscape genetic connectivity and structuring, creating a template on which post-dispersal processes such as local selection, competition, predation and exogenous disturbances operate (Linhart & Grant 1996; Kalisz *et al.* 2001; Nathan & Muller-Landau 2000; Vekemans & Hardy 2004). Within populations, spatial restriction in pollen dispersal may result in correlated and consanguineous mating (Ritland 1989; Muona *et al*. 1991; Hardy *et al*. 2004), which, if seed dispersal is also restricted, will contribute to kinship structure and local inbreeding among naturally regenerated recruits, enhancing local genetic drift and erosion across generations. Under current scenarios of rapid human-mediated landscape change, there is increasing interest to better understand and quantify the effects of restricted pollen dispersal in reproductive biology and spatial genetic structuring (e.g., Koenig & Ashley 2003).

One important theoretical tool for measuring dispersal is the kernel function, defined as the probability density function of dispersal distances from individual plants (Levin & Kerster 1974). The pollen dispersal kernel interacts with the spatial arrangement of individual plants in determining the distribution of mating distances, an interplay with important reproductive implications that has also methodological consequences, since the observed mating distance distribution will generally differ from the normally unobservable kernel function (Robledo-Arnuncio & Austerlitz 2006). Estimating the pollen dispersal kernel has important analytical and modelling applications, such as comparing male dispersal and reproductive success parameters across populations with contrasting individual distributions (Burczyk *et al.* 1996; Oddou-Muratorio *et al.* 2005; Slavov *et al.* 2009), quantifying metapopulation connectivity (Klein *et al.* 2006), and forecasting pollen introgression risk from crops or exotic plantations into natural ecosystems in different demographic settings (e.g. Kuparinen & Schurr 2007).

Genetic paternity analysis is a common technique of choice for fitting pollen dispersal kernels. It requires individual genotypes from a sample of seeds of known maternal origin and from all potential pollen donors within the study area. General methods yield kernel parameter estimates that maximize the likelihood of the observed sample of seed paternal haplotypes, given the Mendelian transition probabilities between pollen donors and seed paternal haplotypes, and given the spatial distribution of pollen donors relative to maternal plants (Adams *et al*. 1992; Burczyk *et al*. 1996; Oddou-Muratorio *et al*. 2005). These methods

do not require a categorical assignment of paternity, and are thus useful for low-resolution genetic assays. During the last years, however, the development of hyper-polymorphic markers is frequently allowing unambiguous paternity assignments, whereby, for every seed, either a single pollen donor is assigned or all potential donors within the study plot are excluded as fathers. Although such level of genetic resolution should facilitate accurate estimation of kernel parameters, a common mistake upon unequivocal paternity assignment is to equate the observed distribution of mating distances and the kernel function, without accounting for the spatial geometry of the population (Robledo-Arnuncio & García 2007).

Properly accounting for spatial population geometry is an essential step towards a better phenomenological description of plant dispersal. A basic and central concern, which has received strikingly low attention in the literature, is how much statistical confidence we can place on paternity data at different distances within the study area, especially at long distances, where we can expect low signal –to-noise ratios. Specific procedures to assess the statistical reliability of paternity assignments for kernel estimation are missing, though they would be most useful for sampling design optimization and formal hypothesis testing. Spatial point pattern analysis techniques (Stoyan & Stoyan 1994; Illian *et al.* 2008), common in other fields of ecology (Wiegand & Moloney 2004), might prove useful for this purpose.

The outcome of unambiguous paternity analyses targeting dispersal kernel fitting comprises the locations of all potential pollen donors within the study area, the locations of selected mother plants, and the paternal origin of seed samples harvested from the mothers. Such data sets can be interpreted as complex marked point patterns (Stoyan & Stoyan 1994), and related spatial analysis methods allow estimation of empirical pollen dispersal kernels and should be able to reveal whether estimated dispersal probabilities at increasing distances are statistically informative, especially with regard to potentially non-random patterns of individual plants. Broadly, randomizing individual identities over all potential donors, while keeping the spatial attributes of the remaining paternity data, provides a null random mating model against which the empirical pollen dispersal kernel can be confronted. Then, simulation envelopes for the null model, constructed via Monte Carlo procedures, allow for formal testing whether the empirical pollen dispersal kernel significantly differs from that expected under random mating, and therefore how informative paternity data is at different distances for kernel fit purposes.

We formalize and apply this approach here to characterize spatial patterns of pollen flow in a population of *Populus nigra* in Germany, where the species is of conservation concern. Our overall objective is to estimate a suitable kernel for modelling pollen movement

in *P. nigra* stands. More specifically, we address the following questions: (1) Does mating occur spatially uncorrelated within the study area? (2) Up to what distance does paternity analysis provide reliable information for kernel fit purposes? Finally, we fit kernel functions to the empirical kernel function resulting from point-pattern analysis and to the observed distribution of mating distances and compare them against results from established maximum-likelihood fitting methods using the competing sources approach (as in Robledo-Arnuncio & García 2007).

Methods

1. STUDY SPECIES AND STUDY SITE

Our study species is the Eurasian black poplar (*Populus nigra* L.), a dioecious and wind pollinated tree with high potential for gene flow over long distances. Black poplar faces two major threats caused by human influence. Firstly, its habitat, the floodplain of rivers, has been reduced by river regulation and intensive utilisation of the agricultural landscape next to the riverbank. This has led to increasing isolation of the fragmented populations. And secondly, its hybrid form *P. x canadensis* (a crossing with the American black poplar *P. deltoides*) is planted along roads and ponds for ornamental purposes. Pollen and ovules are fertile and back crossings with their parents are possible (Bradshaw *et al.* 2000) leading to an unintentional genetic admixture that threatens the species status of *P. nigra*.

Our study site is the floodplain forest of the Eder River next to the city of Fritzlar in Central Germany. Most of our studied trees are standing in a stretch of about 3 km (Fig. 1). Some single trees are located along two smaller rivers parallel to the Eder River in the North and South and in the surrounding landscape that consists mainly of agricultural fields and rural villages.

2. DATA SAMPLING AND GENETIC PATERNITY ANALYSIS

Data collection

In total, we collected leaf material of 331 black poplar trees and measured their geographic position with a differential Leica GS50 positioning system. During the year 2007 we studied the flowering of our trees to differentiate them into males and females, because there is no genetic marker available yet to do that. We counted a total of 197 potential pollen donors, including trees that did not flower during the observation period and that were also defined as potential males. Seeds were harvested directly from the branches of six mother trees during

the year 2006 and 2007. Mother trees were chosen for a representative coverage of the study area and of different local neighbourhoods like dense and open stands (Fig. 1). We collected seeds from different branches of each mother tree to ensure random sampling. To get enough leaf material for genetic fingerprinting, we sowed 200 seeds per mother on Petri dishes and harvested the seedlings after four days.

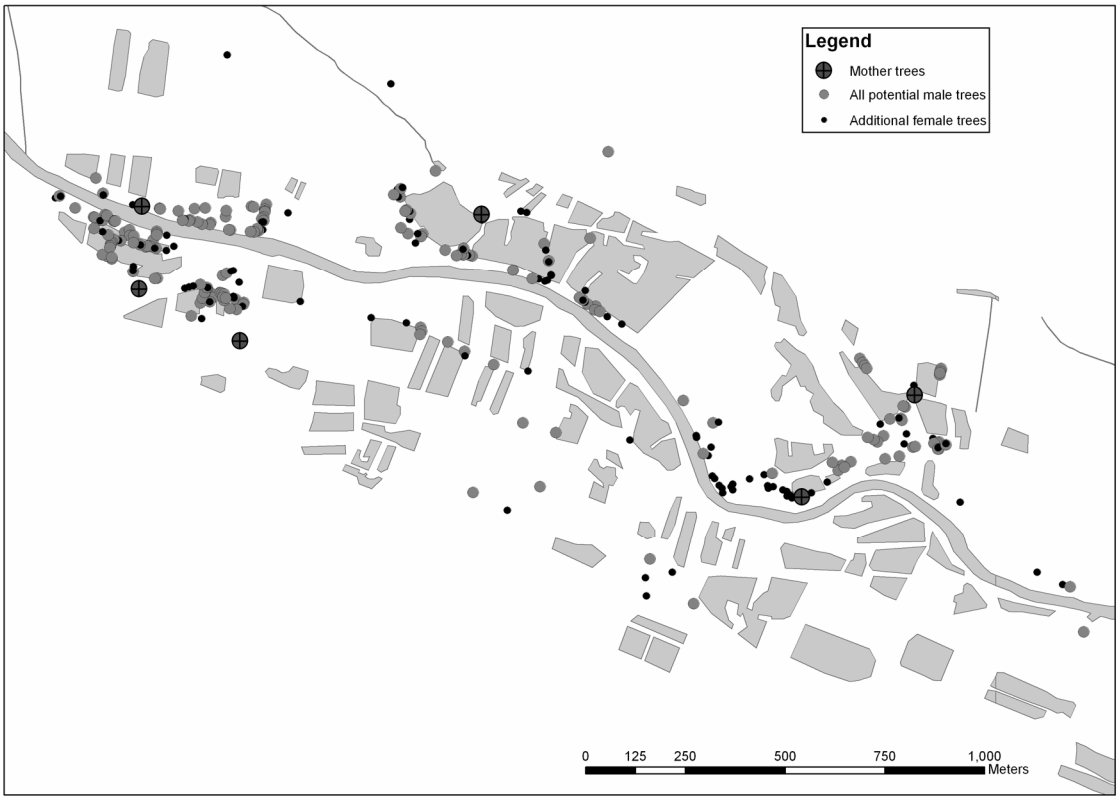


Fig. 1. Spatial distribution of *Populus nigra* male trees (grey dots), sampled mother trees (crossed black disks) and additional female trees (small black dots) along the Eder River.

Microsatellite analysis

All potential pollen donors, mother trees and every individual seedling were genotyped at seven nSSR loci: WPMS05 and WPMS09 (van der Schoot *et al.* 2000), WPMS14, WPMS18 and WPMS20 (Smulders *et al.* 2001) and PMGC14 and PMGC2163 (IPGC, http://www.ornl.gov/sci/ipgc/ssr_resouce.htm). The seven markers are located on different chromosomes (Cervera *et al.* 2001, Gaudet *et al.* 2008) and are therefore unlinked. PCR and fragment electrophoresis protocols are described elsewhere (Rathmacher *et al.* 2009). The combination of the seven loci yielded a very low probability of identity (PI = 1.98×10^{-7}) among individuals within the *P. nigra* stand. This allows for a highly accurate paternity assignment.

Paternity assignment

We conducted paternity assignment using the information of all seven analysed nSSR loci with the computer program Cervus 3.0 (Kalinowski *et al.* 2007). Likelihood ratios for paternity inference are derived using the codominant genetic marker information taking account of typing error and defining a statistic Δ for resolving paternity. Using simulations, the program generates criteria for Δ that permit assignment of paternity to the most likely male with a known level of statistical confidence (Marshall *et al.* 1998). To find the confidence level of paternity several parameter had to be set for the simulations. We simulated 10,000 offspring with 0.01 as the proportion of loci mistyped and 0.8 as the proportion of candidate fathers sampled. We choose 250 individuals as probable candidate fathers. The allele frequencies that were used in the simulation step were calculated based on microsatellite analysis of the adult population. Alleles private to seedlings were added to the frequency data file with p = 0.0001. We used 95 % as strict and 80 % as relaxed confidence level as suggested by Marshall *et al.* 1998.

3. POINT PATTERN ANALYSIS

Mark-correlation functions

Our data set comprises (i) the spatial coordinates of sampled mother trees and of all potential pollen donors in the population, and (ii) based on the paternity analysis, the number of seeds of each mother tree assigned to each pollen donor. Analysis of this data structure allows using methods of marked point patterns which are embedded into the framework of spatial point pattern analysis (Stoyan and Stoyan 1994; Illian *et al.* 2008). The locations of the potential pollen donors (indexed by k) and of the mother trees (indexed by i) represent spatial point patterns, and the marks m_{ki} are the number of seeds of a given mother i fathered by male k.

Mark-correlation functions allow testing if the marks are spatially correlated, conditionally on the spatial locations of male and mother trees. In the standard situation with one pattern and an attached mark [e.g., locations of trees and its size (Illian *et al.* 2008)] mark-correlation functions investigate, for instance, if the mark "size" depends on distance between two points of the pattern. A non-normalized mark correlation function $c_m(r)$ gives for two arbitrary points p and q of the pattern which are distance r apart the expectation of an appropriate test function involving the marks m_p and m_q of the two points. This test function may be for example m_p or $m_p \times m_q$ (Illian *et al.* 2008). In practice, all pairs of points are visited and the average of the test function is calculated for a given distance r. If the mark correlation function $c_m(r)$ is at small distances r smaller that its expectation without spatial

structure of the marks, we have negative correlation which would indicate that nearby trees are smaller than expected by the average size of the trees.

However, our data structure is slightly more complex. We have a focal pattern (i.e., the mother trees which are analogous to "seed traps" in studies of seed dispersal kernels) from which to measure the distance *r* to the donors, and a marked pattern (i.e., the potential donor trees) where the mark m_{ki} is the number of seeds the donor k fathered at a given mother i. To test if the values of the marks m_{ki} depend on the distance *r* to the mother i we therefore used the following summary statistic adapted to our data structure (Fig. 2).

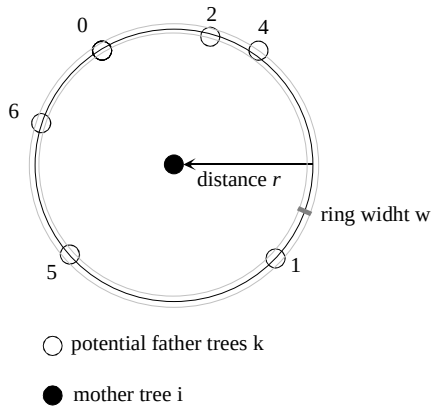


Fig. 2. Illustration of the estimator of the mark correlation function c_{fm}(r). At distance (r-w/2, d+w/2) from the mother tree there are 6 potential father trees which contributed 2, 4, 1, 5, 6, and 0 pollen to the mother tree. Thus, **MarkSumMales**[$R_i^w(r)$] = 2 + 4 + 1 + 5 + 6 + 0 = 18 and **Males**[$R_i^w(r)$] = 6. If mother i would be the representative mother tree, the expected number of seeds fathered by a representative male located distance r away from mother i would yield 18/6 = 3. In practice, the average over all females is taken as estimate of the typical female.

We visited all mother trees i in sequence and counted the marks m_{ki} of all potential donors k located at distance *r* from mother i (= **MarkSumMales**[$R_i^w(r)$]) and the number of potential donors k at distance *r* from the mother tree i (= **Males**[$R_i^w(r)$]), where $R_i^w(r)$ is the ring with radius *r* and width *w* centred in the i^{th} mother. Our summary statistic which is related to the pollen dispersal kernel (see below) is the sum of all marks m_{ki} of potential donors k located at distance *r* from mothers i, divided by the number of potential donors located at distance r from the mothers:

$$\hat{c}_{fm}^{w}(r) = \frac{\sum_{i=1}^{n} \textbf{\textit{MarkSumMales}}\left[R_i^w(r)\right]}{\sum_{i=1}^{n} \textbf{\textit{Males}}\left[R_i^w(r)\right]} \qquad (1)$$

where *n* is the total number of mother trees. Thus, the summary statistic $c_{fm}(r)$ can be interpreted as the average number of seeds donated by male trees located distance *r* away

from a representative mother tree. This function is basically a bivariate r-mark correlation function (Illian *et al.* 2008).

Thus, the summary statistic $c_{fm}(r)$ is the expected number of seeds of a "representative" mother fathered by a "representative" male, given that both are separated by distance r and given the observed seed harvest. The mark correlation function $c_{fm}(r)$ is proportional to the pollen probability density $f(r)$ at distance r from the location of the dispersing father tree which is often called dispersal kernel (e.g., Clark *et al.* (1999)). Following Stoyan and Wagner (2001), the location of a single pollen grain follows in the isotropic case the probability density function $f(r)$: the probability that the pollen grain is in an infinitesimal disc of area dxdy centred at the point (x,y) is $f(r)$dxdy where r is the distance of (x,y) from the location of the pollen dispersing male. The normalization of pollen probability density $f(r)$ is given by

$$1 = \int_0^\infty f(r) 2\pi r dr = 2\pi c \int_0^\infty c_{fm}(r) r dr \,. \qquad (2)$$

For example, when fitting the empirical mark correlation function $c_{fm}(r)$ with an exponential function $a \exp(-r/\alpha)$ the normalizing constant c yields $c = 1/(2\pi a\alpha^2)$. In this way, it is possible to obtain a parametric estimate of the dispersal kernel $f(r)$ by least-square regression of $c_{fm}(r)$ on distance, which accounts for the observed spatial arrangement of both father and mother trees. Note that the pollen probability density $f(r) = c \cdot c_{fm}(r)$ is related to the probability density function $d(r)$ of the random distance of a single pollen from the centre of the tree where $d(r) = 2\pi r f(r)$, thus, $d(r)dr$ is fraction of pollen that are dispersed within a ring of radius r and width dt centred at the dispersing male tree.

Estimators of mark-correlation functions

For estimation of the mark-correlation function we used equation 1 which does not apply edge correction as recommended by Illian *et al.* (2008). We followed the grid-based approach of Wiegand & Moloney (2004) and used a grid size of 5m×5m which guarantees that there was not more than one male tree per cell. This is a sufficiently fine resolution to study our questions. Note that the estimator of the mark-correlation function given in equation 1 involves a decision on the ring width w which basically defined what is regarded as "distance r". Too small ring widths w will produce jagged plots of $c_{fm}(r)$ because not enough points fall within the different rings. We used in all analyses a ring width of 20 cells (= 100m). Note that values of $c_{fm}(r)$ at distance r are somewhat biased for distances $r < w/2$ (Illian *et al.* 2008) because here the distance r is smaller than the ring width w. We therefore used a correction.

We calculated the real area of each ring and determined the real mean radius r_{cor} corresponding to the scale r (see appendix Fig. A1). In all graphs we used r_{cor}.

Test of significance

The empirical summary statistics were contrasted to that of the null model where the marks m_{ki} were randomly shuffled among all potential pollen donators, i.e., there was no spatial structure in the marks (random mating). We used a Monte Carlo approach for construction of simulation envelopes of the null model. Each of the 199 simulations of the point process underlying the null model generates a $c_{fm}(r)$ function and simulation envelopes with an approximate $\alpha = 0.05$ were calculated for the test statistic using its 5^{th} highest and 5^{th} lowest values.

Note that we cannot interpret the simulation envelopes as confidence intervals because we tested the null hypothesis at many scales r simultaneously. This may cause type I error (Stoyan and Stoyan 1994; Diggle 2003; Loosmore & Ford 2006). To test overall departure of the data from the null model without type I error inflation we used a Goodness-of-Fit (GoF) test that collapses the scale-dependent information contained in the test statistics into a single test statistic u_i which represents the total squared deviation between the observed pattern and the theoretical result across the scales of interest. The u_i were calculated for the observed data ($i = 0$) and for the data created by the $i = 1,..199$ simulations of the null model and the rank of u_0 among all u_i is determined. If the rank of u_0 is larger than 190 there is a significant departure from the null model with $\alpha = 0.05$ over a scales of interest. Details can be found in (Diggle 2003, Loosmore & Ford 2006, Illian *et al.* 2008).

4. KERNEL FITS

Based on paternity assignment information obtained with Cervus software, pollen dispersal kernels were estimated in three ways: (i) by direct maximum-likelihood fitting of the observed distribution of mating distances, which does not account for the relative spatial distribution of males and females, (ii) by least squares regression of the empirical estimate of the dispersal kernel using the approach of marked point pattern analyses described above, and, for comparative purposes, (iii) using the competing sources approach (Robledo-Arnuncio & Gil 2005; Robledo-Arnuncio & García 2007). The latter method models the expected proportion of seeds sired by each pollen donor on each maternal progeny as a function of the assumed kernel and the spatial position of the pollen donor relative to the focal female and all

potential pollen donors in the area. Assuming that the vector of observed seed counts for each mother is multinomially distributed with parametric frequencies equal to the expected seed proportions, it yields maximum-likelihood parameter estimates of the assumed kernel function (see Robledo-Arnuncio & García 2007 for details).

We fit several two-dimensional dispersal kernels $f(r; \theta)$, with parameter set θ, yielding the probability of pollen transport per unit area at distance r from the source. First, we considered dispersal kernels from the exponential-power family:

$$f_{ep}(r;a,b) = \frac{b}{2\pi a^2 \Gamma(2/b)} \exp\left(-\left(\frac{r}{a}\right)^b\right). \quad (3)$$

with a and b being the scale and shape parameters, respectively. The average distance of dispersal is given by $\delta = a\ \Gamma(3/b)/\Gamma(2/b)$. The exponential-power distribution includes the Gaussian ($b = 2$) and the exponential ($b = 1$) as special cases. For $b < 1$, the distribution becomes fat-tailed, relative to the exponential. Besides considering the full exponential-power model, in which both a and b were fitted, we also considered a model with fixed $b = 1$, estimating a only, for comparison with abundant previous studies using exponential dispersal kernels.

Second, we considered two-component dispersal kernels. This kind of kernel has been used attempting to describe better the short- and long-distance components of dispersal, under the assumption that they follow different patterns that are not well fitted by simpler probability laws. We used a two-component model comprising two exponential-power functions:

$$f_{ep+ep}(r; a_1,b_1,a_2,b_2,p) = p \cdot f_{ep,1}(r; a_1,b_1) + (1 - p) \cdot f_{ep,2}(r; a_2,b_2) \quad (4)$$

Additionally, we tested a two-component model compound of an exponential-power and a uniform function:

$$f_{ep+U}(r;a,b,c,p) = p \cdot f_{ep}(r;a,b) + (1-p)\left(\frac{1}{\pi c^2}\right) I_{[0,c]}, \quad (5)$$

where c is the upper bound of the uniform distribution and $I_{[0,c]}$ is an indicator function, equalling 1 if $r \leq c$ and zero otherwise. The uniform function yields constant probability of pollen dispersal per unit area in a circular region of radius c around a pollen donor. Note that this does not mean that the distribution of pollen movement distances is uniform, as in Slavov *et al*. (2009), which implies decreasing probability of dispersal *per unit area* with distance, yielding worse fits to our data (results not shown), but rather that a pollen grain has the same probability of moving to any point within the circular region. This compound kernel assumes

that local dispersal follows a predictable pattern while long-distance mating events are governed by highly stochastic processes (e.g. atmospheric turbulences). The mean dispersal distance is given in this case by: $\delta = p[a\ \Gamma(3/b) / \Gamma(2/b)] + (1 - p)(2c/3)$. As before for one-component models, we tested the full model with parameters (a, b, c, p), and then fixed $b = 1$ and estimated (a, c, p) only.

For each of the three estimation methods separately (direct fit, point-pattern analysis and competing sources), we used Akaike's information criterion (AIC) to select the most parsimonious pollen dispersal model among those tested. For least squares regression (point pattern analysis), assuming normally distributed errors, we used AIC = $n\log(\mathrm{RSS}/n) + 2K$, where RSS is the residual sum of squares, n is the number of observations and K is the number of estimated parameters (including the intercept and RSS/n). For maximum-likelihood fits, we used AIC = $-2\log(L(\hat{\theta})) + 2K$, with $L(\hat{\theta})$ being the value of the likelihood function at its estimated maximum.

Results

1. GENETIC PATERNITY ANALYSIS

Out of a total sample of 2,599 seedlings, we could assign 1,535 (395 in 2006; 1,140 in 2007) seedlings to their fathers with a probability of more than 80 % based on LOD scores (50 % of the assigned seedlings were even above 95 %). The genetic paternity analysis revealed long distance dispersal events from fathers up to 7,500m away from the mother tree in the year 2007. However, for most mother trees, the most distant father is about 2,000m away. Overall, 56 % of effective pollen has its origin in a distance of less than 500m, 74 % less than 1,000m and 95 % less than 2,000m (Fig. 3). The relative spatial position of individual trees in the population determined a non-uniform distribution of potential pollen dispersal distances, visually similar to the distribution of realized mating distances at intermediate and long distances (Fig. 3). At short distances, however, there was an apparent excess of mating events, relative to the potential distribution.

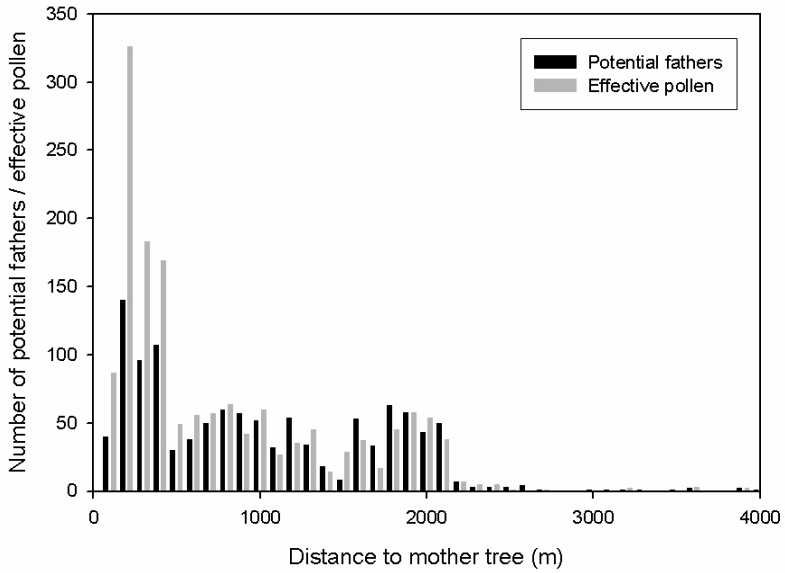


Fig. 3. Distribution of potential pollen dispersal distances and observed mating distances for *Populus nigra* along the Eder River in Central Germany.

Because of the low number of potential males and fathered seeds in distances of more than 2,000m, the point pattern analysis as well as the kernel fit was done for distances up to 2,000m (Fig. 3).

2. POINT PATTERN ANALYSIS

Question 1: Does mating occur spatially uncorrelated within the study area?

After accounting for the non-uniform distribution of potential dispersal distances, the point pattern analysis showed that empirical pollen dispersal kernels declined with increasing *r* and that there was an excess of nearby mating but fewer mating events occurred at larger distances (Fig. 4). As expected, the joined 2006 and 2007 data set (Fig. 4C) shows this tendency better than the estimates based only on the 2006 or 2007 data sets (Figs. 4A, B). Thus, the null model of random mating does not describe the data well: the empirical kernel function is especially at short distances above the simulation envelopes and at larger distances close to or below the simulation envelopes (Fig. 4).

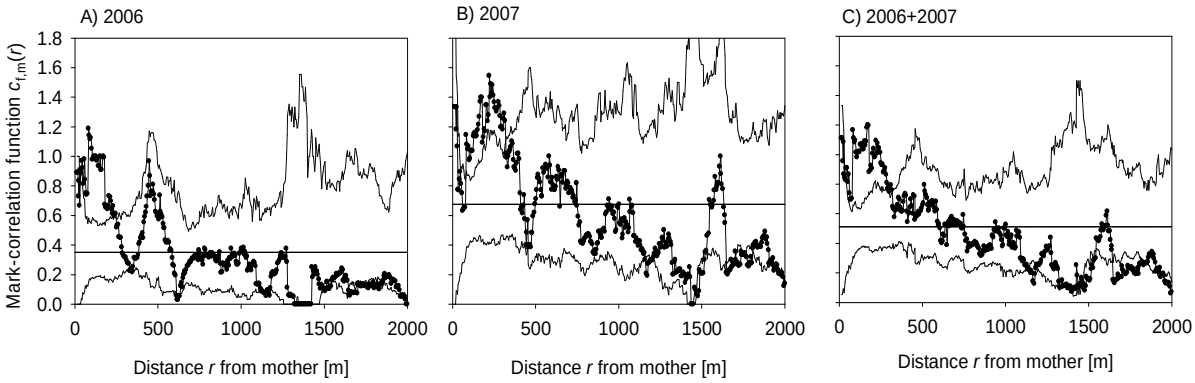


Fig. 4. Results of the mark-correlation analyses on *Populus nigra* paternity data. (A) The mark m_{ki} attached to potential donor k was the number of seeds of mother i fathered by donor k. In this case the mark-correlation function $c_{fm}(r)$ gives the expected number of seeds a potential donor would father at a representative mother tree, given it is located at distance r from the mother. Closed circles: mark-correlation function $c_{fm}(r)$; horizontal line: expectation under the null model where the marks were randomly shuffled over all potential fathers; lines: simulation envelopes being the 5th lowest and highest values taken form 199 simulations of the null model. (B) Same as A) but for 2007 data (C) Same as A) but for joined 2006 and 2007 data.

To formalize these observations, we applied the GoF test. The GoF test for distances 0-1000m yielded for the 2006, 2007, and joined 2006 and 2007 data ranks of 199, 194, and 200, respectively (Table 1). This indicates a significant departure from the null model of random mating, obtained by shuffling of marks of the 2006 and joined 2006 and 2007 data (Table 1). However, the distance interval 0-2000m was only weakly significant for the 2006+2007 data (Table 1).

Table 1. Rank of the Goodness-of-Fit (GoF) test for different distance intervals. The estimate of the kernel function is at a 0.05 error level significantly different from the null model of random shuffling of the father trees if rank > 190 and at the 0.01 error level if rank > 198.

Data set	0-500m	0-1000m	0-2000m	500m-2000m	1000m-2000m	1500m-2000m
2006	199**	199**	193*	174	180	191*
2007	197*	194*	184	147	160	173
2006+2007	200**	200**	198*	188	192*	198*

Question 2: Up to what distance does paternity analysis provide reliable information for kernel fit purposes?

Although the empirical dispersal kernels showed a clearly decreasing shape (i.e., an excess of nearby mating at distances smaller than 600m and a deficit of mating at larger distances (Fig. 4C) our approach using Monte Carlo simulations of the random mating null model allowed us to test if the dispersal kernel at larger distances differs significantly from the null model. First indications are given by the simulation envelopes: the empirical kernel function is above the simulation envelopes between 30m and 240m in 2006 (Fig. 4A), between 80m and 310m in 2007 (Fig. 4B), and between 40m and 310m for the combined 2006+2007 data set (Fig. 4C). Additionally, the empirical kernel is for some larger distances below the simulation envelopes (Figs. 4A, C). This let us suspect that there are significantly less effective pollination events at distances larger than 1500m than expected under random mating.

To avoid problems of type I error inflation associated with the simulation envelopes we tested for departures from the null model at different distance intervals (Table 1). The data sets for individual years provide no evidence for departures from random mating at larger scales except a weak effect for 1500-2000m for the 2006 data set. However, the combined data set revealed a significant difference in the kernel function for 1000-2000m and 1500-2000m from the null model (Table 1). Thus, the tail of the distribution can be distinguished from a null model without spatial dependence, but because stochastic effects are strong (as manifested by the wide simulation envelopes) the uncertainty on the detailed shape of the decline is large. The reason for this is that the number of potential donors decreased at our study design strongly with distance from the mother (Fig. 3) and therefore we observe that stochastic effects are strong at larger distances r. This makes the simulation envelopes wide. To further illustrate this high uncertainty at larger distance we show in Figure A2 three examples of mark correlation functions simulated by the null model.

The general shape of the distribution functions $c_{fm}(r)$ were consistent between the two study years but we observed nevertheless differences in detail: the number of pollen in 2007 was higher than in 2006 (cf. Figs. 4A and 4B) and in 2007 pollen arrived more frequently from male trees located further away than in 2006.

Question 3: Is it possible to fit a dispersal kernel to the point pattern data and if yes is this kernel comparable with a kernel using maximum-likelihood fitting methods?

The fitting of a dispersal kernel to the results of the mark-correlation function provides estimates consistent with results from the competing sources model (Fig. 5).

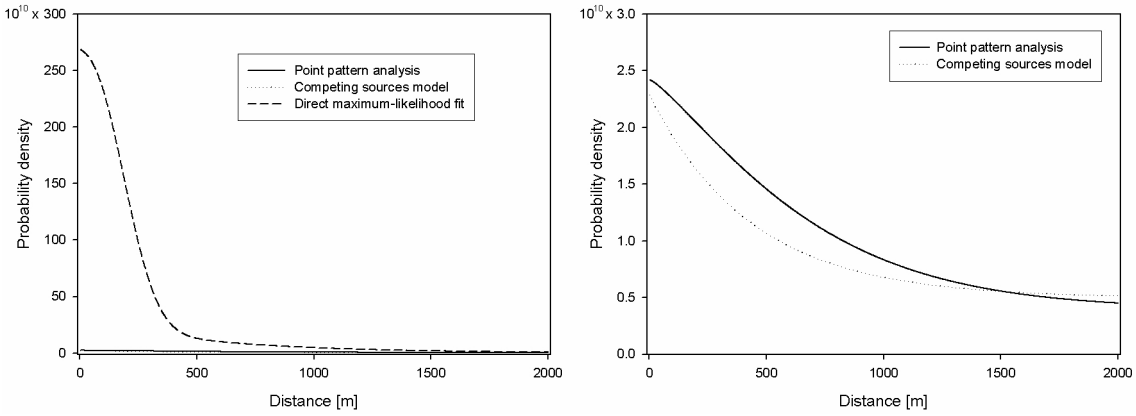


Fig. 5. Estimated pollen dispersal kernels for *Populus nigra*. The left panel shows the best fit models obtained with the three used methods: point pattern analysis (PP), competing sources model (CS) and direct maximum-likelihood fit of the observed mating distance distribution (dML). See text for details about the assumed dispersal models. The right panel compares PP and CS results only (Y-axis rescaled for clarity).

The mean dispersal distance of the kernel function providing the best fit is 1567m for the point pattern data, while it is 1275m for the competing sources method. By contrast, the best direct fit to the observed mating distance distribution yielded a kernel with much larger probability density at short distances (Fig. 5) and a substantially lower mean dispersal distance (704m; Table 2), as expected from the excess of potential fathers over shorter distances (Fig. 3).

Table 2. Pollen dispersal kernel estimates for Populus nigra using three different methods: point pattern analysis (PP), the competing sources model (CS) and direct maximum-likelihood fit of the observed mating distance distribution (dML). See text for details about the assumed dispersal models. The best model for each data set is shown in bold, as indicated by Akaike's information criterion (AIC). The negative log-likelihood (for CS and dML methods) and residual sum of squares (for PP method) are also shown for each fit.

Method	Model	Parameter estimates			*mean* (m)	- Log-likelihood/ RSS	AIC
		Component 1	*Component* 2	*p*			
PP	*Exponential*	a=994.0929 b=1 (fixed)	-	-	1988.2	5.4557	-1823.7
	Exp-pow	a=927.3593 b=0.8792	-	-	2436.7	5.3767	-1827.8
	2C *Exppow*	a_1=671.9442 b_1=1.5203	a_2=2697.9714 b_2=1999.8711	0.2172	1568.8	4.8144	-1868.3
	2C Exp + Unif	a=700.4653 b=1 (fixed)	c=2101.4696	0.4904	1401.0	5.1216	-1846.2
	2C Exppow + Unif	***a*=671.9430 *b*=1.5203**	***c*=2694.5863**	**0.2175**	**1567.1**	**4.8144**	**-1870.3**
CS	*Exponential*	a=1247.3165 b=1 (fixed)	-	-	2494.6	15262.3	30526.6
	Exp-pow	a=866.8769 b=0.6857	-	-	4580.1	15256.9	30517.8
	2C *Exppow*	a_1=513.1915 b_1=1.1965	a_2=2097.5551 b_2=1424.8318	0.2171	1259.1	15210.6	30431.2
	2C Exp + Unif	***a*=447.9898 *b*=1 (fixed)**	***c*=2096.8934**	**0.2455**	**1274.7**	**15209.7**	**30425.5**
	2C Exppow + Unif	a=502.8345 b=1.1806	c=2096.9853	0.2144	1260.4	15209.0	30426.0
dML	*Exponential*	a=346.1 b=1 (fixed)	-	-	692.2	23048.7	46099.5
	Exp-pow	a=40.5333 b=0.5152	-	-	695.5	22901.5	45806.9
	2C *Exp-pow*	a_1=133.4227 b_1=0.9319	a_2=2103.0626 b_2=68.4221	0.6383	699.7	22666.6	45343.2
	2C Exp + Unif	***a*=152.5994 *b*=1 (fixed)**	***c*=2095.9943**	**0.6346**	**704.3**	**22657.8**	**45321.6**
	2C Exppow + Unif	a=149.5671 b=0.9905	c=2095.9943	0.6347	703.9	22657.8	45323.5

The best fit for the point pattern data is obtained with a compound kernel comprising an exponential power and a uniform function. The model with two exponential power functions yielded a very similar fit (ΔAIC = 2), since in fact the parameter estimates of the first exponential power function were virtually the same (i.e., $a \approx a_1$ and $b \approx b_1$; Table 2) and the second-component exponential power, by having a very large shape parameter b_2 = 1999.9 and a scale parameter of a_2 = 2698.0, yields virtually identical probabilities than the uniform

function with c = 2694.6. For the other two methods, the competing sources model and the direct fit, a compound exponential-uniform function is more suitable, following Akaike's information criterion, although the compound exponential power-uniform function has a very similar AIC value ($\Delta AIC \leq 2$; Table 2). The fact that the best fits for all three estimation methods corresponded to compound kernels including a uniform probability function indicates the highly stochastic nature of observed mating events at long distances.

Discussion

We analyzed the spatial structure of pollen dispersal and estimated pollen dispersal kernels based on results of a genetic paternity analysis. Marked point pattern analysis (Stoyan & Stoyan 1994; Wiegand & Moloney 2004; Illian *et al.* 2008) produces an empirical (non-parametric) estimate of the kernel function. This approach takes the spatial structure of males and females directly into account. More importantly, it allows confrontation of the empirical kernel function with kernel functions generated by a null model that assumes no spatial structure in mating, i.e., each donor within the study area has the same probability to father seeds at any mother tree within the study area. This null model corresponds to a truncated uniform pollen dispersal kernel, which is not implausible *a priori* for a wind pollinated species such as our study species *P. nigra* and should thus be formally tested Some studies have used non-parametric tests (e.g. Mann-Whitney-Wilcoxon) to check whether the observed distribution of mating distances significantly departs from the distribution of potential dispersal distances i.e., whether mating occurs at random or not (see Streiff *et al.* 1999). The novelty of the marked point pattern analysis developed here in the context of dispersal kernel estimation is that it provides a powerful permutation test that allows testing and quantifying this departure over different distance classes.

The use of marked point pattern analyses showed a clear and significant decrease of the probability of effective pollen dispersal with distance and could confirm a highly significant excess of mating events at short distances (say < 500m) and a somewhat weaker, but significant shortage of mating events at larger distances (say > 1500m-2000m). The clear distance dependency in the empirical dispersal kernel up to few hundred meters from the mother outlines the strong influence of local spatial structure on offspring paternity. These results fit to the observations of Rathmacher *et al.* (accepted) that isolation by distance is present in our population.

We observed that the duration and time of flowering of individual trees differed between the years (unpublished data). This is in line with findings of limited pollination events in other *P.*

nigra populations because of its flowering phenology (Vanden Broeck *et al.* 2003). Having two years of data would improve our kernel estimates. Indeed, the combined 2006 and 2007 data set evened out such yearly fluctuations and provided a better "average" description of the underlying process, enabling a more powerful test of random-mating departure than any of the two single-year patterns (Table 1). The marked point pattern analysis could reveal a range up to 240m in 2006 and 310m in 2007 to show excess of mating at shorter distances. These differences between the years might be due to the larger number of assigned seeds in 2007 and different meteorological conditions during the flowering seasons of the two consecutive years.

Based on the point pattern analysis of the two-year paternity data, the best kernel fit was obtained with a two component distribution comprising an exponential power and a uniform probability function. This compound kernel is consistent with the raw point pattern results in that it reflects a predictable decay of mating probability up to a certain distance (i.e., the exponential power component) and a high uncertainty on the shape of the kernel's tail (i.e., the truncated uniform distribution), reflecting the small number of mating events at large distances, which were (weakly) significantly below random mating expectations.

The fitting of a kernel function for the combined 2006 and 2007 data with the competing sources model, which has been shown to yield reliable estimates under varied demographic and dispersal conditions (Robledo-Arnuncio & García 2007; Jones & Muller-Landau 2008), yielded results consistent with that of the point pattern analysis. As was the case for the point pattern analysis, model selection using the competing sources approach favoured clearly two component kernels comprising a uniform component (ΔAIC of the two one-component kernels relative to the selected two-component kernels was larger than 90; Table 2).). The scale and shape of the best models obtained with both the point pattern and competing sources methods were comparable, as indicated by the estimated mean dispersal distances (1567 vs. 1275 m), shape parameter of the first component *b* (1.52 vs. 1.00), and weighting factor (1 - *p*) for the second (larger distance) uniform component (0.78 vs. 0.75).

However, the question remains whether long distance dispersal occurs truly without distance dependence, being governed by highly stochastic processes such as atmospheric turbulences (Nathan *et al.* 2002), or whether it responds to a continuously decreasing function with distance, only that we are unable to detect it empirically. Clearly, estimation noise at large distances hampers the accurate measurement of the kernel's tail, and thus it could be that the uniform probability function was selected as the second component of the most

parsimonious kernel simply because the data at larger distances did not carry sufficient information to detect a distance dependence pattern.

Confrontation of the empirical dispersal kernel to those derived from Monte Carlo simulations of the null model (that maintained our data structure except its spatial structure) yielded somewhat frustrating results. Figure A2 shows some examples of simulated dispersal kernels which arise through randomization of the data. They cover a wide range from increasing to decreasing pollen probability with distance, yielding uncomfortably wide simulation envelopes. The consequence of this result is that studies using statistical fitting may substantially underestimate the uncertainty in the estimates of dispersal kernels at larger scales. Doubling the sample size (by combining the 2006 and 2007 data sets) eased this problem, but much larger sample sizes are needed to produce reliable results at larger scales. This is a general dilemma in dispersal kernel studies (see Kot *et al.* 1996), which deserves further theoretical consideration. On the other hand, the spatial pattern of our species was extremely clustered which constrained the occurrence of intermediate and larger pollen dispersal distances thereby somewhat aggravating the uncertainty at the tail of the dispersal kernels.

It is common practice to obtain "kernel" estimates by directly fitting the observed mating distance distribution (Okubo & Levin 1989; Darmency *et al.* 2009). This approach ignores the fact that the distribution of potential dispersal distances (i.e., potential male-female distances) is not uniform and the competition between the different pollen donors is ignored. Our results show that the two used methods, the marked point pattern analysis and the competing sources model, which take the geometry of the mother and donor trees explicitly into account, yield similar kernel parameter and mean dispersal distance estimates. Slavov *et al.* (2009) accounted for spatial structure as well and found a similar dispersal kernel for *P. trichocarpa* in the US. However, as the fitting of the data included distance classes up to 10,000m, the mean dispersal distance was five times bigger than our results. By contrast, the simple direct fitting method that ignores population geometry yielded a substantially smaller mean dispersal distance estimate. The latter is most likely downwardly biased, because spatial aggregation of donors around mother trees produces a bias towards smaller dispersal distances if no correction is applied that takes the total number of potential donors at a given distance into account.

With our study species *P. nigra* we could reveal distance dependent gene flow in the Eder River floodplain. The possibility to assess significant departures from the null model of random mating (or any other appropriate null model) is an important gain of the marked point

pattern approach to account for spatial structure of male and female trees during pollen dispersal, especially over long distances. The fitted kernel function seems suitable for further studies regarding the simulation of pollen mediated gene flow in wind pollinated trees.

Acknowledgements

M.N. was funded by the Federal Ministry of Education and Research, Germany (Grant ID 0313285J).

References

Adams, W.T., Griffin, A.R. & Moran, G.F. (1992) Using paternity analysis to measure effective pollen dispersal in plant populations. *American Naturalist*, **140**, 762-780.

Austerlitz, F., Dick, C.W., Dutech, C., Klein, E.K., Oddou-Muratorio, S., Smouse, P.E. & Sork, V.L. (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937-954.

Burczyk, J., Adams, W.T. & Shimizu, J.Y. (1996) Mating patterns and pollen dispersal in a natural knobcone pine (*Pinus attenuata* Lemmon.) stand. *Heredity*, **77**, 251-260.

Bradshaw, H.D., Ceulemanns, R., Davis, J. & Stettler, R. (2000) Emerging model systems in Plant Biology: Poplar (Populus) as a model forest tree. *Journal of Plant Growth Regulation*, **19**, 306–313

Cervera, M.T., Storme, V., Ivens, B., Gusmao, J., Liu, B.H., Hostyn, V., Van Slycken, J., Van Montagu, M. & Boerjan, W. (2001) Dense genetic linkage maps of three Populus species (*Populus deltoides, P-nigra* and *P-trichocarpa*) based on AFLP and microsatellite markers. *Genetics*, **158**, 787-809.

Clark, J.S., Silman, S., Kern, R., Macklin, E. & HilleRisLambers, J. (1999) Seed dispersal near and far: Patterns across temperate and tropical forests. *Ecology*, **80**, 1475-1494.

Darmency, H., Klein, E.K., De Garanbe, T.G., Gouyon, P.H., Richard-Molard, M., & Muchembled, C. (2009) Pollen dispersal in sugar beet production fields. *Theoretical and Applied Genetics* **118**, 1083-1092.

Diggle P. J. (2003) Statistical Analysis of Spatial Point Pattern. Hodder Arnold Publication.

Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Mugnozza, G.S. & Sabatti, M. (2008) Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. *Tree Genetics & Genomes*, **4**, 25-36

Hardy, O.J., González-Martínez, S.C., Colas, B., Fréville, H., Mignot, A. & Olivieri, I. (2004) Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. *Genetics*, **168**,1601-1614.

Illian, J., Penttinen, A., Stoyan, H. & Stoyan, D. (2008) Statistical Analysis and Modelling of Spatial Point Patterns. Wiley, Chichester.

Jones, F.A. & Muller-Landau, H.C. (2008) Measuring long-distance seed dispersal in complex natural environments: an evaluation and integration of classical and genetic methods. *Journal of Ecology*, **96**, 642-652.

Jones, M.E., Shepherd, M., Henry, R. & Delves, A. (2008) Pollen flow in *Eucalyptus grandis* determined by paternity analysis using microsatellite markers. *Tree Genetics & Genomes*, **4**, 37-47.

Kalinowski, S.T., Taper, M.L. & Marshall, T.C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, **16**, 1099-1106.

Kalisz, S., Nason, J.D., Hanzawa, F.M. & Tonsor, S.J. (2001) Spatial population genetic structure in *Trillium grandiflorum*: The roles of dispersal, mating, history, and selection. *Evolution*, **55**, 1560-1568.

Klein, E.K., Lavigne, C. & Gouyon, P.H. (2006) Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. *BMC Ecology*, **6**, 3.

Koenig, W.D. & Ashley, M.V. (2003) Is pollen limited? The answer is blowin' in the wind. *Trends in Ecology & Evolution*, **18**, 157-159.

Kot, M., Lewis, M.A., & vandenDriessche, P. (1996) Dispersal data and the spread of invading organisms. *Ecology*, **77**, 2027-2042.

Kuparinen, A. & Schurr, F.M. (2007) A flexible modelling framework linking the spatio-temporal dynamics of plant genotypes and populations: application to gene flow from transgenic forests. *Ecological Modelling*, **202**, 476-486.

Levin, D.A. & Kerster, H.W. (1974) Gene flow in seed plants. *Evolutionary Biology*, **7**, 139-220.

Linhart, Y.B. & Grant, M.C. (1996) Evolutionary significance of local genetic differentiation in plants. *Annual Review of Ecology and Systematics*, **27**, 237-277.

Loosmore, N.B. & Ford, E.D. (2006) Statistical inference using the G or K point pattern spatial statistics. *Ecology*, **87**, 1925-1931.

Marshall, T.C., Slate, J., Kruuk, L.E.B. & Pemberton, J.M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, **7**, 639-655.

Muona, O., Moran, G.F. & Bell, J.C. (1991) Hierarchical patterns of correlated mating in *Acacia melanoxylon*. *Genetics*, **127**, 619-626.

Nathan, R., Katul, G.G., Horn, H.S., Thomas, S.M., Oren, R., Avissar, R., Pacala, S.W., & Levin, S.A. (2002) Mechanisms of long-distance dispersal of seeds by wind. *Nature*, **418**, 409-413.

Oddou-Muratorio, S., Klein, E.K. & Austerlitz, F. (2005) Pollen flow in the wildservice tree, *Sorbus torminalis* (L.) Crantz. II. Pollen dispersal and heterogeneity in mating success inferred from parent-offspring analysis. *Molecular Ecology*, **14**, 4441-4452.

OECD. Consensus document on the biology of Populus L. (poplars), Organisation for Economic Co-operation and Development, Series on Harmonization of Regulatory Oversight in Biotechnology No. 16, Paris, 2000.

Okubo, A. & Levin, S.A. (1989) A theoretical framework for data analysis of wind dispersal of seeds and pollen. *Ecology*, **70**, 329-338.

Rathmacher, G., Niggemann, M., Wypukol, H., Gebhardt, K., Ziegenhagen, B. & Bialozyt, R. (2009) Allelic ladders and reference genotypes for a rigorous standardization of poplar microsatellite data. *Trees – Structure and Function*, **23**, 573-583.

Rathmacher, G., Niggemann, M., Köhnen, M., Ziegenhagen, B. & Bialozyt, R. (accepted). Short-distance gene flow in *Populus nigra* L. accounts for small-scale spatial genetic structures - implications for in-situ conservation measures. Conservation Genetics.

Ripley B.D. (1981) Spatial Statistics. Wiley.

Ritland, K. 1989. Correlated matings in the partial selfer *Mimulus guttatus*. *Evolution*, **43**, 848-859.

Robledo-Arnuncio, J.J. & Gil, L. (2005) Patterns of pollen dispersal in a small population of *Pinus sylvestris* L. revealed by total-exclusion paternity analysis. *Heredity*, **94**,13-22.

Robledo-Arnuncio, J.J. & Austerlitz, F. (2006) Pollen dispersal in spatially aggregated populations. *American Naturalist*, **168**, 500-511.

Robledo-Arnuncio, J.J. & García, C. (2007) Estimation of the seed dispersal kernel from exact identification of source plants. *Molecular Ecology*, **16**, 5089-5109.

Slavov, G.T., Leonardi, S., Burczyk, J., Adams, W.T., Strauss, S.H. & Difazio, S.P. (2009) Extensive pollen flow in two ecologically contrasting populations of *Populus trichocarpa*. *Molecular Ecology*, **18**, 357-373.

Smulders, M.J.M., van der Schoot, J., Arens, P. & Vosman, B. (2001) Trinucleotide repeat microsatellite markers for black poplar (*Populus nigra* L.). *Molecular Ecology Notes*, **1**, 188-190.

Stoyan, D. & Stoyan, H. (1994) Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics. Wiley, Chichester.

Stoyan, D. & Wagner, S. (2001) Estimating the fruit dispersion of anemochorous forest trees. *Ecological Modelling*, **145**, 35-47.

Streiff, R., Ducousso, A., Lexer, C., Steinkellner, H., Glössl, J. & Kremer, A. (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand from *Quercus robur* and *Quercus petraea* (Matt.) Liebl. *Molecular Ecology*, **8**, 831-841

Tabbener, H.E. & Cottrell, J.E. (2003) The use of PCR based DNA markers to study the paternity of poplar seedlings. *Forest Ecology and Management*, **179**, 363-376.

Vekemans, X. & Hardy, O.J. (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology*, **13**, 921-935.

Van der Schoot, J., Pospiskova, M., Vosman, B. & Smulders, M.J.M. (2000) Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.) populations. *Theoretical and Applied Genetics*, **101**, 317-322

Vanden Broeck, A., Cox, K., Quataert, P., Van Bockstaele, E. & Van Slycken, J. (2003) Flowering phenology of *Populus nigra* L., *P. nigra cv. italica* and *P. x canadensis* Moench. and the potential for natural hybridisation in Belgium. *Silvae Genetica*, **52**, 280-283.

Wiegand, T. & Moloney, K.A. (2004) Rings, circles, and null-models for point pattern analysis in ecology. *Oikos*, **104**, 209-229.

Wiegand, T., Gunatilleke, S., Gunatilleke, N. & Okuda, T. (2007) Analyzing the spatial structure of a Sri Lankan tree species with multiple scales of clustering. *Ecology*, **88**, 3088-3102.

Appendix

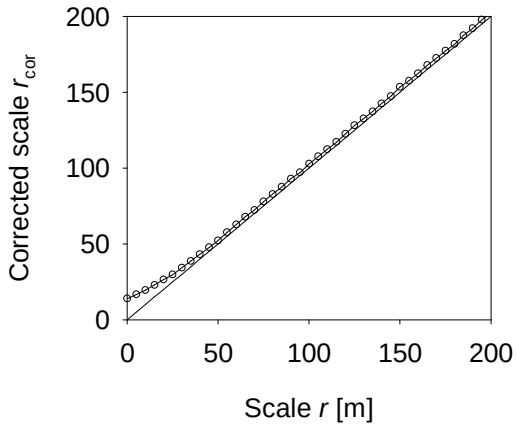


Fig. A1. Small bias in scale r caused by the ring width w = 100m. The bias arises because the ring is defined as ($r - w/2$, $r + w/2$) and for $r < w/2$ = 50m only the outer half of the ring exist (see fig. A1) so that the average distance form inside the ring to the origin (= r_{cor}) is larger than r.

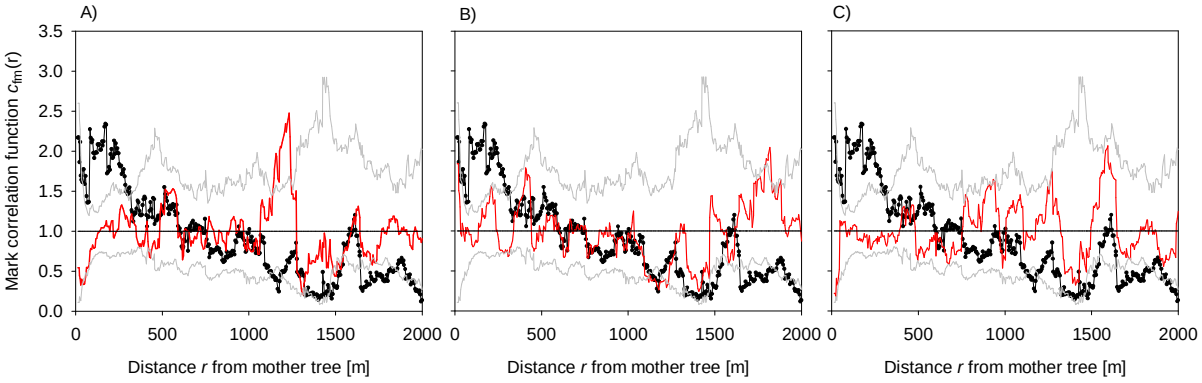


Fig. A2. Three different examples (A, B and C) of kernel functions for the combined 2006 and 2007 *Populus nigra* paternity data that can arise through the null model, i.e., random assignation of the identity of pollen donors. The observed non-normalized kernel function (black dots), the simulation envelopes (grey lines) and examples of non-normalized kernel function arising from the null model (red). Horizontal line: expectation under the null model where the marks were randomly shuffled over all potential fathers.

Paper 3

Pollen competition and restricted gene flow between *Populus nigra* and its hybrid form *Populus x canadensis*

Manuscript

with G. Rathmacher and R. Bialozyt

Abstract

The extent and the consequences of interspecies gene flow are in the focus of evolutionary biologists as well as conservation ecologists.

Crossings between the black poplar *Populus nigra* and its hybrid form *P. x canadensis* has been examined in experiments and in natural populations. However, the strength of introgression and the pollen competition between natural and hybrid poplar pollen is hardly measured.

Therefore we created a model for pollen mediated gene flow in a real landscape. We conducted simulation studies to explain the discrepancy between rates of interspecies gene flow and the proportion of hybrid trees in the vicinity of sampled mother trees in the floodplain of the Eder River in Western Germany. We tested three common probability density functions as well as a pollen dispersal function that has been successfully fitted to the intra-population gene flow of our studied black poplar population.

We could show that it is necessary to include a barrier for the crossing of hybrid males with the female trees to picture the hybrid proportions in the seed set that we found in our studied population. The result was independent of the used dispersal functions and our barrier has comparable values with previous barriers of hand pollination experiments.

Introduction

Exchange of individuals among populations is essential for species survival in a stochastic and changing environment. Otherwise, in an evolutionary perspective gene flow among individuals is the first step in creation of new genotypes. The new combinations of successful alleles are the basis for evolution and adaptation. However, exchange and crossings between distant and different adapted populations of one species could also lead to outbreeding depression (Waser and Price 1989). The same process is possible at the scale of discrete species (Bleeker et al. 2007). Hybridisation of close related species is common in some plant species (e.g. Orchids), whereas in others it is restricted by not overlapping habitats or differentiated reproduction mechanisms (Ellstrand et al. 1996). Human changing of habitat and the introduction of former alien species enables them to bridge the natural ecological and geographic barriers threatening the species status of native species (Levin et al. 1996). In some cases humans even actively cross species to benefit from the often better hybrid forms especially in agriculture and silviculture. In Western Europe the most common hybrid tree in the landscape is *Populus x canadensis*, a crossing between the American *Populus deltoides* and the European black poplar *Populus nigra*. This fast growing hybrid tree is used for timber as well as for ornamental purposes for the last 300 years.

Because the back crossings of the hybrid with its natural parent threaten the pure species status of the already endangered natural floodplain species *P. nigra*, several studies have been conducted to detect natural introgression. Some of them found open pollinated seeds or offspring of hybrid parents (Pospíšková and Šálková 2006; Smulders et al. 2008; Ziegenhagen et al. 2008; Rathmacher et al., in prep.), others could not (Imbert and Lefevre 2003; Fossati et al. 2003; Tabbener and Cottrell 2003). This is due to the often small sample size making the results vulnerable to stochastic effects and due to pollen competition. It has been shown by hand pollination experiments that hybrid pollen has a disadvantage becoming active on female poplars (Benetka et al. 2002; Vanden Broeck et al. 2003a). These findings indicate the existence of a pre- or postzygotic barrier in artificial crossings but were not tested in natural stands so far. Pollen competition in the context of hybridisation is known for different plant species (reviewed in Rieseberg & Carney 1998). In *Populus* species a prezygotic barrier based on the growth of the pollen tub has been observed (Villar et al. 1993). Differences in the flower phenology of *P. nigra* and *P. x canadensis* have also shown a possible restriction in interspecies gene flow (Vanden Broeck et al. 2003b). However, a total overlap of the flowering time is possible due to special climate conditions in some years.

Modelling the pollen dispersal using density functions is a common approach in gene flow studies (Austerlitz et al. 2004). The results of paternity analyses make it possible to estimate the pollen dispersal curve. However, simulation studies have yet not been used to calculate the pollen competition respectively the barrier of a hybrid crossing with its natural parent. In contrast to hand pollination experiments the necessary data for a modelling approach, namely coordination of potential fathers and effective pollen respectively seeds assigned to its parents, are often collected in parentage analyses routinely.

In this study we look into the discrepancy between the genetically revealed proportion of hybrid fathers in sampled seeds from a natural population of *P. nigra* and the number and proportion of potential hybrid fathers in the vicinity of the mother trees. To estimate the incoming pollen mixture at the mother tree comparable to artificial hand pollination, simulations using different dispersal kernels were conducted to calculate the amount and composition of the incoming pollen.

We ask the following question: Is it possible to picture the genetic results by introducing a barrier for the hybrids reducing the success of hybrid pollen in the female trees?

Methods

1. Study site and study species

Our study site is an area of about 90km² in Central Germany (Fig. 1). The centre is represented by the floodplain of the Eder River. The riparian forest consists of softwood and hardwood trees dominated by willows and black poplars. [The surrounding landscape consists mainly of agricultural fields and rural villages. There are two smaller rivers parallel to the Eder River in the North and South.]

Our study species are the endangered European black poplar *Populus nigra* and its hybrid form *P. x canadensis* (a crossing with the American black poplar *Populus deltoides*). The hybrids are mostly planted along roads and ponds for ornamental purposes. Pollen and ovules are fertile and back crossings with their parents are possible and leading to fertile offspring. Poplars are dioecious and wind pollinated trees with high potential for gene flow over long distances.

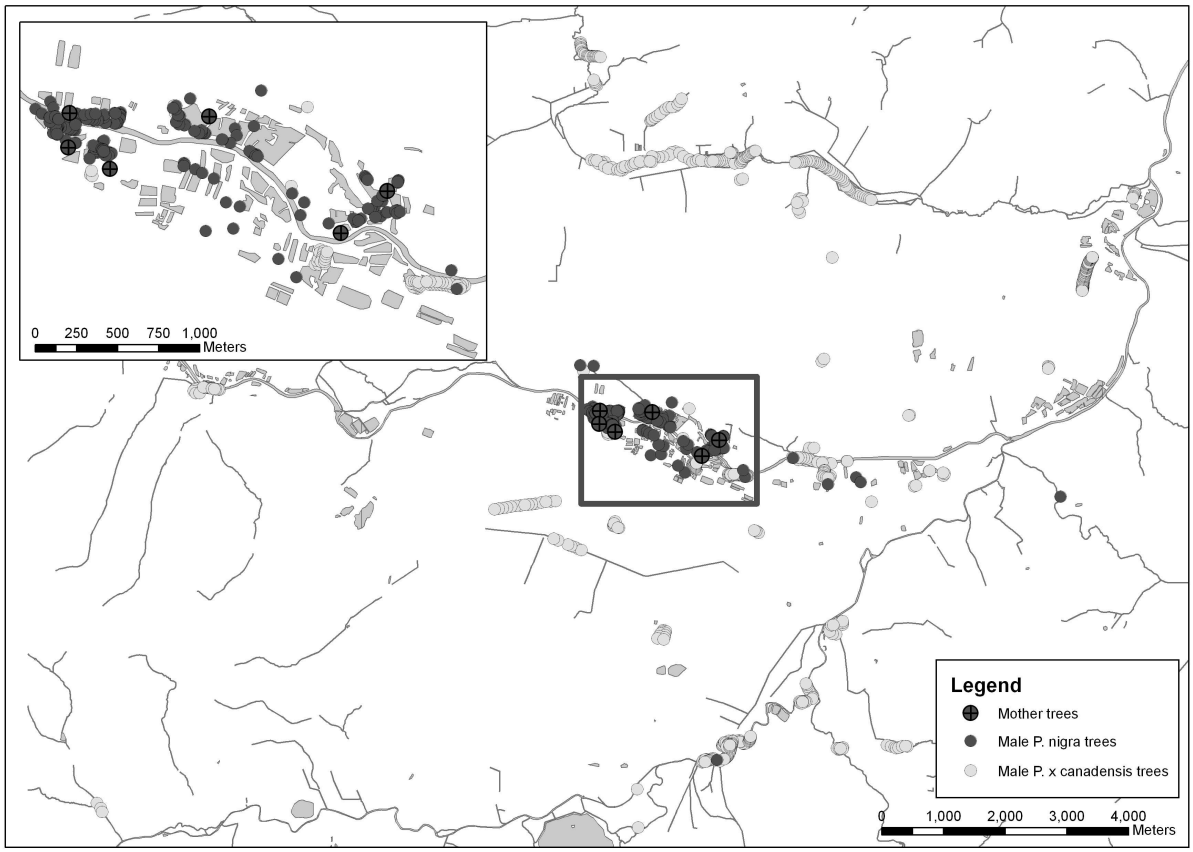


Figure 1. The spatial pattern of male *Populus nigra* trees (dark disks), *Populus x canadensis* trees (light disks) and of the analyzed mother trees (dark crossed disks). The small picture is an enlargement of the core of the study area.

2. Data sampling and introgression rate

Data collection

We recorded 989 adult poplar trees and assigned them genetically to one of the two species *P. nigra* (N, 331) and *P. x canadensis* (DN, 658). The number of hybrid trees is probably even higher because of additional not sampled plantations in the surrounding. In 2006 and 2007, we collected 1,533 seeds from the branches of six female trees (2006: 387 seeds from N mothers, 149 from DN mothers; 2007: 708 seeds from N mothers, 289 from DN mothers).

Molecular analyses

The seeds were analysed genetically following Rathmacher et al. (2009). The proportion of male hybrids in the seeds of *P. nigra* and *P. x canadensis* mothers were calculated. This was done by counting the amount of *P. deltoides* specific alleles using the four diagnostic and unlinked markers WPMS09 (van der Schoot et al. 2000), WPMS18 (Smulders et al. 2001), PMGC14 and PMGC2163 (IPGC SSR Resource). Afterwards, a correction term was applied on the actual proportion of hybrid fathers, because the real number will be underestimated due to Medelian segregation (see Rathmacher et al., in prep).

3. Simulation

We created a spatially explicit simulation model able to disperse pollen from all male trees to all female trees. To verify the robustness of the model the simulations were conducted with different dispersal kernels. We used standard kernel functions of the normal, 2Dt and exponential power families (see Austerlitz et al. 2004) as well as a two component exponential power function fitted to distances of revealed father trees by a genetic paternity analysis (Niggemann et al., submitted). Sensitivity analyses were performed to assess the importance of the standard deviation σ of the kernel functions. No differences between the flight characteristics of *P. nigra* and hybrid pollen are assumed. We used a parameter *z* to reduce the chance of the hybrid pollen to become effective on female trees subsuming zygotic as well as phenological barriers. We picked 1000 seeds per mother tree at random to calculate the proportion of seeds within the sample fathered by male hybrids.

The simulated proportion was compared with the real data by a Bray-Curtis similarity test. The mother trees were the samples.

Results

Without including a barrier in the simulation of hybrid pollen dispersal it was not possible to obtain the proportion of seeds fathered by hybrid trees that we found in our harvested seeds in 2006 and 2007 (Table 1). In both years a barrier *z* of 10^{-1} for the standard kernels and of 10^{-2} for the fitted kernel provides the highest similarity with the natural findings in our studied trees (Fig. 2). Comparing only the results of the 2Dt function using different values for the parameter σ, the stability and robustness of the detected barrier becomes clear. A barrier *z* of 10^{-1} provides the highest similarity when the parameter $\sigma > 100$ (Fig. 3).

Table 1. The proportion of seeds fathered by hybrid trees in the seed set in 2006 and 2007 and the results of the used dispersal kernels for a barrier $z = 0.1$ (Gaussian, 2Dt, Exp-pow) and $z = 0.01$ (Fitted kernel)

in %	1	17	83	146	169	228
2006	6.7	1.0	3.0	0.0	0.0	0.0
2007	3.1	0.0	1.5	1.5	0.0	1.1
Gaussian	3.5	7.5	0.5	0.5	1.0	0.7
2Dt	6.4	6.7	2.4	2.2	2.0	3.1
Exp-pow	5.2	6.9	0.6	0.3	0.7	1.8
Fitted kernel	1.5	1.1	1.3	0.3	1.3	0.7

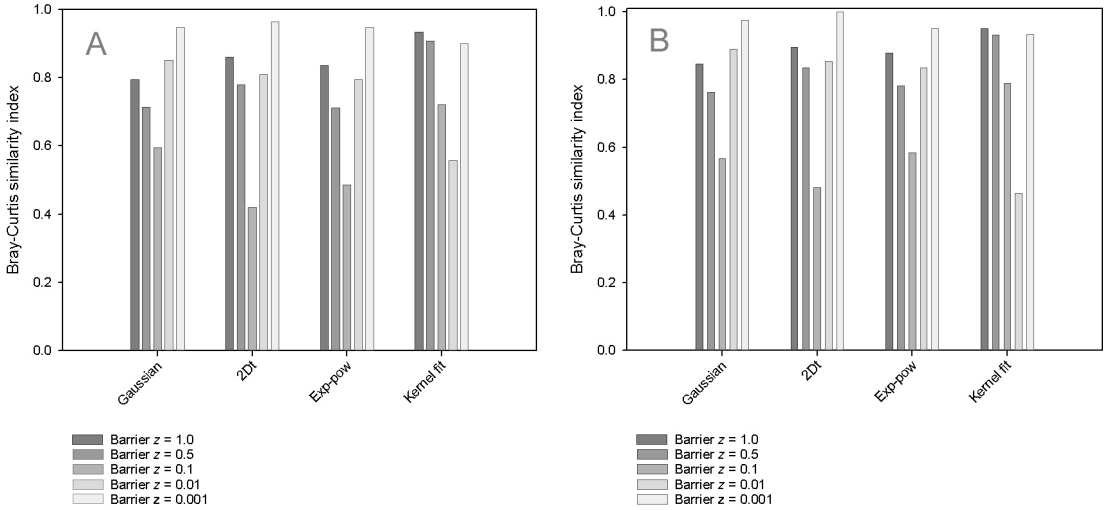


Figure 2. The result of the Bray-Curtis similarity test for different dispersal kernels in (A) 2006 and (B) 2007. Note, low index values indicate high accordance of simulated and real data.

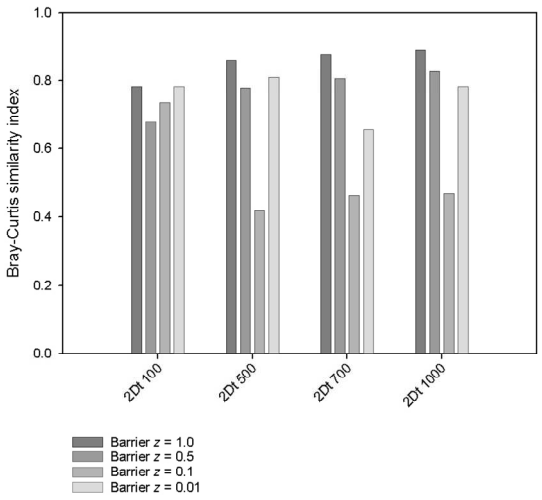


Figure 3. The results of the Bray-Curtis similarity test for the 2Dt function with σ=100, σ=500, σ=700 and σ=1000 for barrier z=1.0, z=0.5, z=0.1 and z=0.01. Note, low index values indicate high accordance of simulated and real data.

Discussion

Pollen competition and the disadvantage of *P. x canadensis* pollen has been demonstrated by hand pollination experiments in *P. nigra*. However, this natural barrier has not been proven in a real landscape during open pollination.

On the basis of an extensive data set of assigned seeds our model could show, that there has to be a barrier for the crossing between the two poplar species *P. nigra* and *P. x canadensis*. The small proportion of hybrid fathers within the sampled seeds could not be explained by the proportion of male hybrid trees standing in the vicinity of the mother trees. According to this, simulating pollen dispersal using no barrier (z=1) always leads to higher differences between

simulated and realised proportions in the seed set than introducing a barrier. This barrier could be the result of differences in the flowering phenology of black poplars and its hybrids. However, in 2006 the two species have been flowering synchronously (Niggemann et al. 2006) and therefore a barrier because of none or limited overlapping flowering periods seems not to be the case.

Therefore a zygotic barrier may be responsible for the differences. Villar et al. 1993 have found a prezygotic barrier in the *Populus* genus. In controlled crossings the pollen of *P. nigra* has always an advantage over pollen of *P. x canadensis* males both when the female is *P. x canadensis* (Rajora 1989) or *P. nigra* (Benetka et al. 2002, Vanden Broeck et al. 2003a). Comparing the proportion of hybrid pollen used in the controlled crossings with the resulting proportion of hybrid fathers in the seeds the data of Vanden Broeck et al. (2003a) provides a barrier or reduction factor of 0.076 and 0.091. Our results of the simulation study with barriers of 0.01 and 0.1 confirm this range. However, our results were not achieved by controlled hand pollination experiments but from a natural *P. nigra* population in two consecutive years. The consistence between the years makes it clear that the barrier is stable and reliable.

The differences between the standard and fitted kernels based on the fact that the fitted kernels were created for the whole study area and therefore are extremely fat-tailed to explain the occurrence of distant black poplar fathers (see Niggemann et al., submitted). In this case, they result in increased pollen loads from the hybrid stands and plantations of the surrounding of the *P. nigra* population comparing to the standard kernels with limited long distance dispersal. Additional, still unknown poplar stands may increase the amount of hybrid pollen leading to an even higher barrier, but the limited and decreasing influence of male trees with increasing distance to the female tree enables only slight variations. Nevertheless, there has to be a barrier regardless of its exact value.

As *P. x canadensis* is already in the focus of genetic modification, crossings between the hybrid and its natural parent not only threatens the species status of *P. nigra* but also enables the introduction of genetic modified alleles into the wild. Although the contamination is limited because of the barrier, it is not impossible even in populations with more than 300 *P. nigra* trees in close vicinity as we could show. In the context of an increase in the number and size of *P. x canadensis* plantations to satisfy the demand for renewable primary products, the proportion of 1:100 of natural black poplars to hybrids is not a distant thread but already the case especially for single female black poplars. Therefore the identification of a barrier

should not make us feel save but let us think about restrictions in the establishment of new plantations of *P. x canadensis* in the close vicinity of *P. nigra* populations.

This is the first time that the advantage of *P. nigra* pollen over *P. x canadensis* pollen is shown in a natural stand of *P. nigra*. We have to make clear that this is only possible by simulation studies because even if the amount of pollen at a female is measured it would not be sufficient as the proportion of hybrid pollen is still unknown in the measured sample as it is not possible to differentiate the pollen.

References

Austerlitz, F., Dick, C.W., Dutech, C., Klein, E.K., Oddou-Muratorio, S., Smouse, P.E., and V.L. Sork (2004). Using genetic markers to estimate the pollen dispersal curve. Molecular Ecology 13: 937-954.

Benetka, V., Vackova, K., Bartakova, I., Pospiskova, M., and M. Rasl (2002). Introgression in black poplar (*Populus nigra* L. ssp. *P. nigra*) and its transmission. Journal of Forest Science 48: 115-120.

Bleeker, W., Schmitz, U., and M. Ristow (2007). Interspecific hybridisation between alien and native plant species in Germany and its consequences for native biodiversity. Biological Conservation 137: 248-253.

Ellstrand, N.C., Whitkus, R., and L.H. Rieseberg (1996). Distribution of spontaneous plant hybrids. Proc. Natl. Acad. Sci. USA 93: 5090-5093.

Fossati, T., Grassi, F., Sala, F., and S. Castiglione (2003). Molecular analysis of natural populations of Populus nigra L. intermingled with cultivated hybrids. Molecular Ecology 12: 2033-2043.

Imbert, E., and F. Lefevre (2003). Dispersal and gene flow of Populus nigra (Salicaceae) along a dynamic river system. Journal of Ecology 91: 447-456.

Levin, D.A., Francisco-Ortega, J., and R.K. Jansen (1996). Hybridization and the extinction of rare plant species. Conservation Biology 10: 10-16.

Niggemann, M., Rathmacher, G., and R. Bialozyt (2006). The risk of introgression of foreign genes in Populus spec. – Differences in the flowering phenology of *Populus nigra* and *P.* × *canadensis*. In: Hoffmeister, T., Diekmann, M. (eds.) Proceedings of the GfÖ 36, p. 91.

Niggemann, M., Wiegand, T., Robledo-Arnuncio, J.J., and R. Bialozyt (submitted). Marked point pattern analysis on genetic paternity data for estimation and uncertainty assessment of pollen dispersal kernels.

OECD. Consensus document on the biology of Populus L. (poplars), Organisation for Economic Co-operation and Development, Series on Harmonization of Regulatory Oversight in Biotechnology No. 16, Paris, 2000.

Pospíšková, M., and I. Šálková (2006). Population structure and parental analysis of black poplar along the Morava River. Canadian Journal of Forest Research 36: 1067-1076.

Rajora, O.P. (1989). Pollen competition among *Populus deltoides* Marsh., *P. nigra* L. and *P. maximoviczii* Henry in fertilizing *P. deltoides* ovules and siring its seed. Sex. Plant Reprod. 2: 90-96.

Rathmacher, G., Niggemann, M., Wypukol, H., Gebhardt, K., Ziegenhagen, B., and R. Bialozyt (2009) Allelic ladders and reference genotypes for a rigorous standardization of poplar microsatellite data. Trees – Structure and Function 23: 573-583.

Rathmacher, G., Niggemann, M., Ziegenhagen, B. and R. Bialozyt (in prep.). Subtle invasion of *Populus deltoides* genes into endangered *P. nigra* L. populations – challenges for quantitative diagnostic methods.

Rieseberg, L.H., and S.E. Carney (1998). Plant hybridization. New Phytologist. 140: 599-624.

Smulders, M.J.M., van der Schoot, J., Arens, P., and B. Vosman (2001). Trinucleotide repeat microsatellite markers for black poplar (*Populus nigra* L.). Molecular Ecology Notes 1:188-190.

Smulders, M.J.M., Beringen, R., Volosyanchuk, R., Vanden Broeck, A., van der Schoot, J., Arens, P., and B. Vosman (2008). Natural hybridisation between *Populus nigra* L. and *P. x canadensis* Moench. Hybrid offspring competes for niches along the Rhine river in the Netherlands. Tree Genetics and Genomes 4: 663-675.

Stettler, R.F., Zsuffa, L. and R. Wu (1996). The role of hybridization in the genetic manipulation of *Populus* in: Stettler, R.F., Bradshaw, H.D., Heimann, P.E. and T.M. Hinckley (eds.) Biology of *Populus*. NRC Research Press, Ottawa, pp 87-112.

Tabbener, H.E., and J.E. Cottrell (2003). The use of PCR based DNA markers to study the paternity of poplar seedlings. Forest Ecology and Management 179: 363-376.

Van der Schoot, J., Pospiskova, M., Vosman, B., and M.J.M. Smulders (2000). Development and characterization of microsatellite markers in black poplar (*Populus nigra* L.). Theor Appl Genet 101:317-322.

Vanden Broeck, A., Quataert, P., Roldán-Ruiz, I., Van Bockstaele, E., and J. Van Slycken (2003a). Pollen competition in *P. nigra* females revealed by microsatellite markers. Journal of Forest Genetics. 10: 219–227.

Vanden Broeck, A., Cox, K., Quataert, P., Van Bockstaele, E. and J. Van Slycken (2003b). Flowering phenology of *Populus nigra* L., *P. nigra cv. italica* and *P. x canadensis* Moench. and the potential for natural hybridisation in Belgium. Silvae Genetica 52, 280-283.

Villar, M., Gaget, M., Rougier, M. and C. Dumas (1993). Pollen-pistil interactions in *Populus*: Beta-Galactosidase activity associated with pollentube growth in the crosses *P. nigra* x *P. nigra* and *P. nigra* x *P. alba*. Sex. Plant Reprod. 6: 249-256.

Ziegenhagen, B., Gneuss, S., Rathmacher, G., Leyer, I., Bialozyt, R., Heinze, B., and S. Liepelt (2008). A fast and simple genetic survey reveals the spread of poplar hybrids at a natural Elbe river site. Conservation Genetics 9: 373-379.

APPENDIX

Appendix A

Conceptual work for a new wind dispersal kernel

The Gaussian diffusion-deposition model

There are different kinds of distribution functions commonly used for the dispersal of pollen and seeds (see Austerlitz et al. 2004). However, all of these functions (Normal, exponential, exponential power, Weibull or Student t et al.) based on the Gaussian distribution function and do not include wind direction or wind speed. It has been shown by simulations that poplar pollen is able to reach sites as far as 30 km away from the source (Bialozyt, personal communication). Additionally, a preferred wind direction resulted in completely different wind flags and therefore different pollination probabilities in the studied poplar populations.
This led to the search for a wind dependent dispersal function for the pollen flow of the studied poplar species. The solution seemed to be the Gaussian diffusion-deposition model of Overcamp (1976). It was used for the deposition of fine and heavy particles of industrial emission. The probability *P* of a particle to reach a position x_1, x_2, x_3 is described as
Equation 1:

$$\rho(x_1, x_2, x_3) = \frac{E}{2 \cdot Pi \cdot \sigma_2(x_1) \cdot \sigma_3(x_1) \cdot u} \cdot e^{-\left[\frac{x_2^2}{2 \cdot \sigma_2^2(x_1)} + \frac{(x_3 - h)^2}{2 \cdot \sigma_3^2(x_1)}\right]}$$

with *E* = amount of particles emitted by the source, *h* = height of the source, *u* = constant wind flow in the direction x_1, and σ_2 and σ_3 as distance and wind dependent functions (equation 2).
Equation 2:

$$\sigma_2(x_1) = F \cdot x_1^{\,f}$$
$$\sigma_3(x_1) = G \cdot x_1^{\,g}$$

The standard deviation σ_2 represents the variation in x_2 direction and σ_3 in x_3 direction. The values for the parameters *F*, *f*, *G* and *g* are constant for a specific wind stability class (Table 1).

Table 1: Stability classes A to F and the corresponding values for *F*, *f*, *G*, and *g* based on Klug/ Manier.

Stabilitätsklasse	F	f	G	g
A	0.469	0.903	0.017	1.380
B	0.306	0.885	0.072	1.021
C	0.230	0.855	0.076	0.879
D	0.219	0.764	0.140	0.727
E	0.237	0.691	0.217	0.610
F	0.273	0.594	0.262	0.500

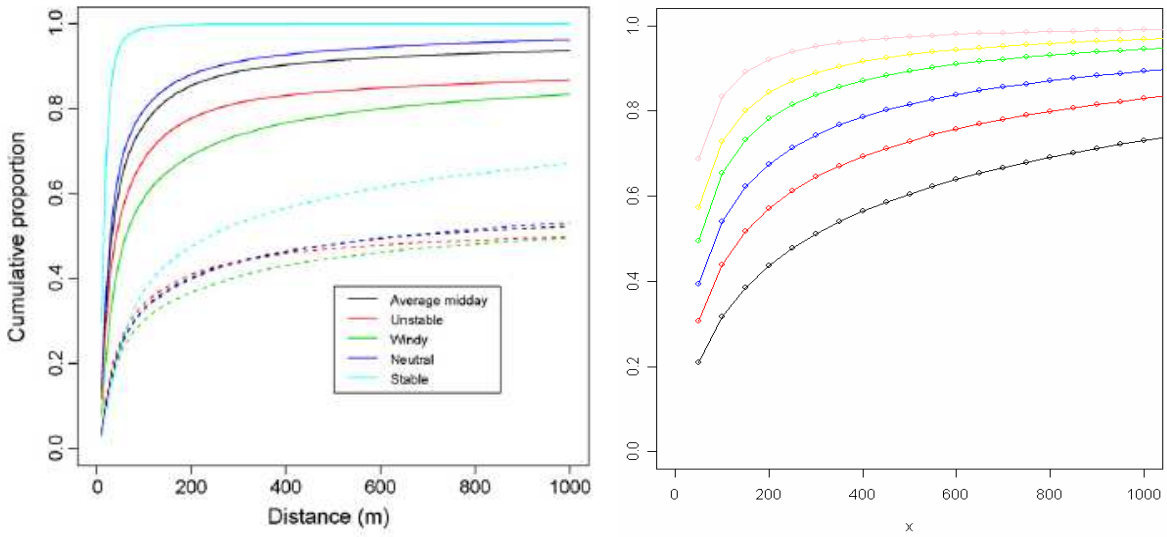
The wind stability classes of Klug/ Manier based on the stability of the atmospheric layer: A – very unstable, B – unstable, C – slightly unstable, D – neutral, E – stable, F – very stable. The wind stability depends basically on the wind strength and the cloudiness. It has to be mentioned that the exact values for *F*, *f*, *G* and *g* are the result of a dispersion experiment in the Upper Rhine valley during the year 1985. Accurately defined amounts of a Tracer were included in the emission of a power plant in Karlsruhe. At 20 sites in a distance up to 60 km the concentration was measured again. The experiment has the following problems:

1. The number of sites where the concentration was measured was limited.
2. The wind direction and strength was neither spatially nor in time constant.
3. Basic assumptions of equation 1 were not fulfilled.

Therefore, it is necessary to bear in mind that there are limitations of the values in table 1. Nevertheless, these values are used in the German TA Luft for the official allowance of the construction of industrial facilities in the contest of air pollution.

Modification of the model

We calculated the cumulative proportion of the pollen with increasing distances (Fig. 1) to compare the results of equation 1 (in the following only called M1) with the mechanistic model of Kuparinen et al. (2007).



Kuparinen et al. 2007 M1

Figure 1: Cumulative proportion of the pollen with increasing distance from the source

The order of the curves from unstable (lowermost) to stable (topmost) is the same in both models. In the mechanistic model there is a set of curves for light (lowermost) and heavy

(topmost) particles. This was not possible in M1, because equation 1 has no parameter for terminal velocity respectively the weight of a particle or pollen. The slope of the curve in Fig 1b is not as considerable as in Fig 1a for the mechanistic model. In M1 the pollen only get onto the ground and stop the dispersal process because of the Brownian movement. The point is that equation 1 is still a diffusion-deposition model.

Therefore we tried three different approaches to account for the terminal velocity of pollen.

Calculation of the σ-values by simulation

The values for *F*, *G*, *f* and *g* based on the results of the dispersion experiment and are therefore specific for the weather conditions and particle size during the experiment. If we want to include the weather conditions during pollen release in our study area and the size and weight of poplar pollen, we had to conduct our own experiment. Because this was not possible, we used a mechanistic model (Kuparinen et al. 2007) to simulate the pollen dispersal (Fig. 2).

Out of the mechanistic model, where we included wind stability and pollen terminal velocity, we received the number of pollen ρ at different distances in x_1 (100, 500, 1000, 2000, 5000) and x_2 (0, 100, 500) direction.

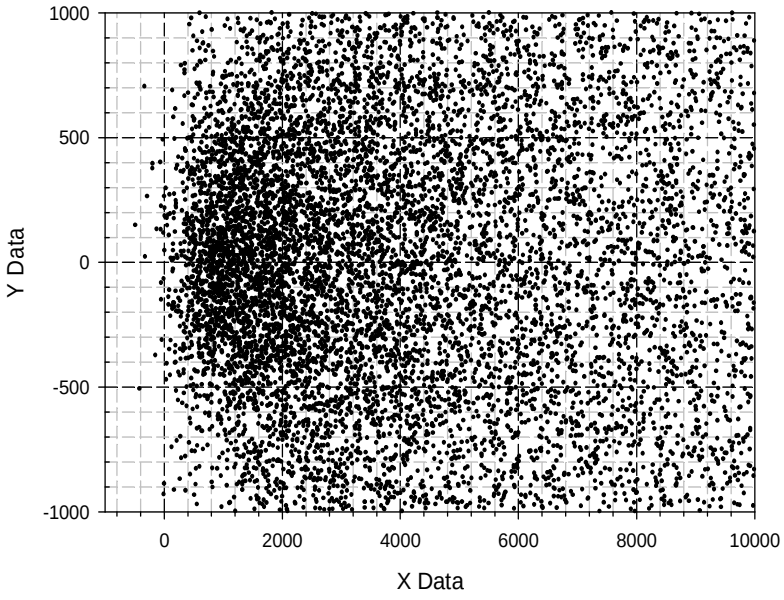


Figure 2: The pollen cloud of the mechanistic model using the local weather conditions and pollen terminal velocity.

Because the values for σ_2 and σ_3 are independent of the x_2 direction (see equation 2), we calculated the values for σ_2 and σ_3 out of equation 1 by using the number of pollen ρ and constant x_1. This was done by a simulation because equation 1 is not solvable analytically.

The results were pairs of σ_2 and σ_3 for each pair of x_1 and x_2 (Fig. 3).

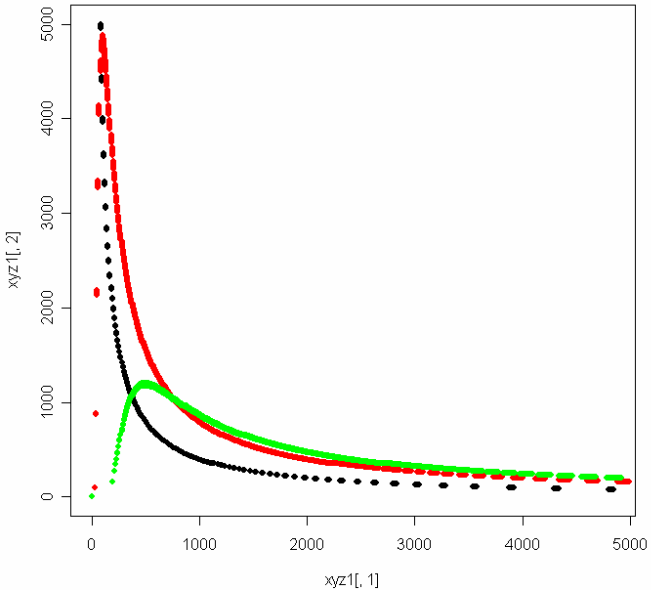


Figure 3: The results of the simulations for all possible σ_2 and σ_3 in the x_1 direction 500m and x_2=0 (black), x_2=100 (red) and x_2=500 (green).

Because there was no reasonable combination of σ_2 and σ_3 for all x_2 directions, it seems to be impossible to calculate *F*, *G*, *f* and *g* using this approach.

Calculation of the σ-values by its standard deviation

Because σ_2 represents the standard deviation of the pollen in x_2 direction within the pollen cloud, it should be possible to calculate the standard deviation out of the mechanistic model directly. This was done by counting the number of pollen within a section with a width of 500m in x_2 direction (Fig. 4).

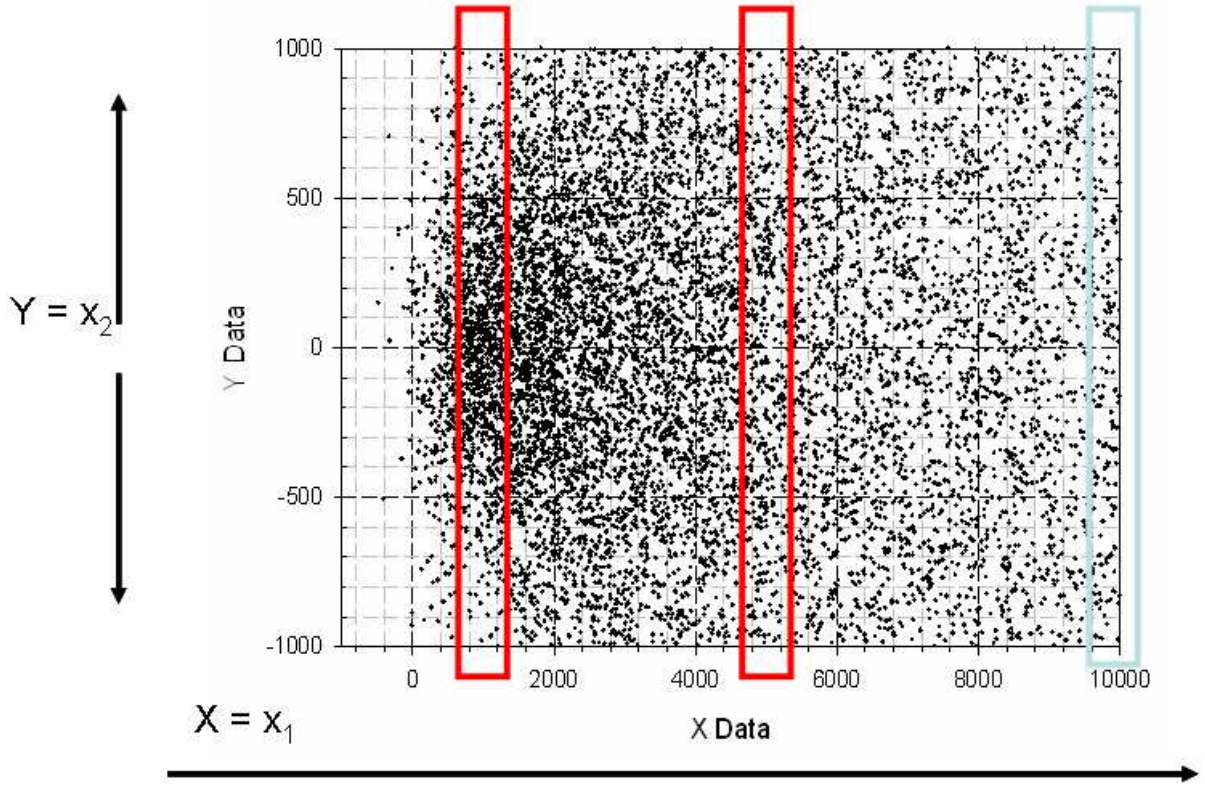


Figure 4: The pollen cloud of the mechanistic model and three sections to calculate σ_2.

The calculation of F and f out of equation 2 using the standard deviation σ_2 for the different sections (see Table 2) was done graphically (Fig. 5).

Table 2: The amount of pollen grains in different sections and the standard deviation σ_2

x_1	250	1000	3000	5000	10000
Pollen	14038	46523	35543	26126	14396
σ_2	433.7	501.03	893.38	1151.64	1508.31

The intercept with the x_2 axis is F and the slope of the regression line is f.

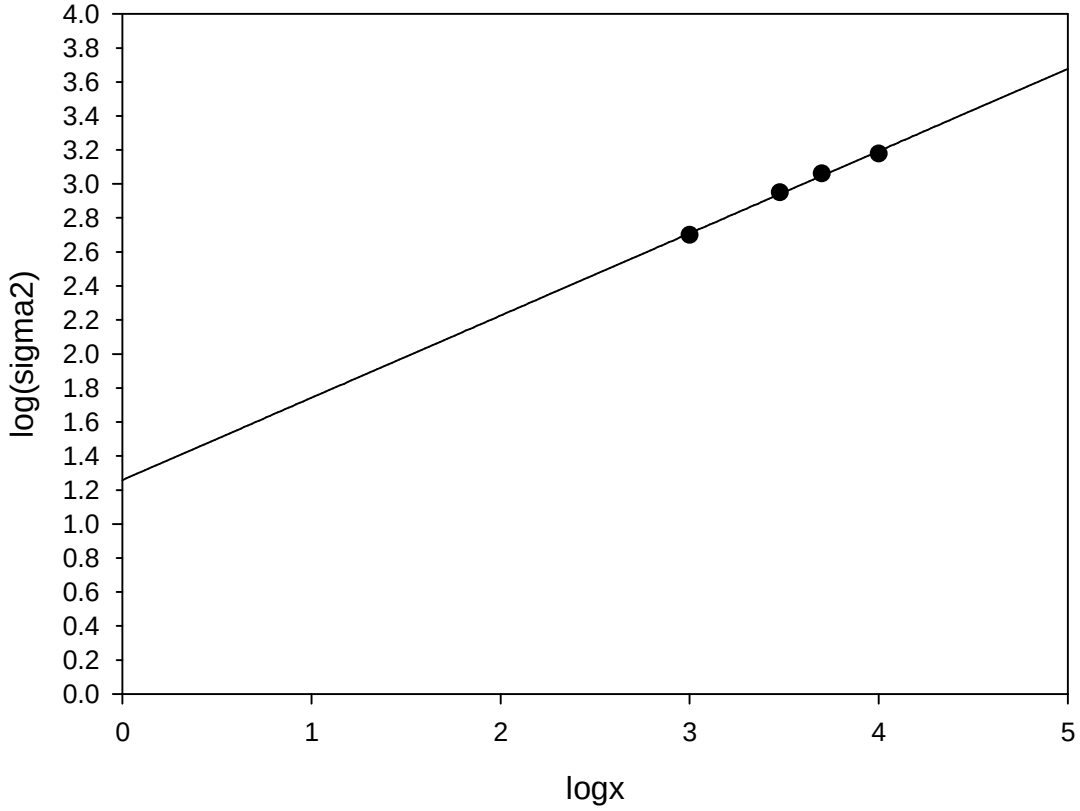


Figure 5: Plotted equation 2 with x_1 and σ_2 on a logarithmic scale to determine F and f.

Because there is only a low number of pollen for small x_1 in the mechanistic model, even little changes in the slope of the regression line will result in a considerable variation in the intercept and therefore in F based on the logarithmic scaling.

Consequently, this approach to determine the σ values using the mechanistic model that included wind stability and terminal velocity is not useful.

Integration of the terminal velocity

Because the calculation of specific σ values that include the terminal velocity was not possible, we introduced the terminal velocity directly in equation 1.

Equation 3:

$$\rho(x_1, x_2, x_3) = \frac{E}{2 \cdot Pi \cdot \sigma_2(x_1) \cdot \sigma_3(x_1) \cdot u} \cdot e^{-\left[\frac{x_2^2}{2 \cdot \sigma_2^2(x_1)} + \frac{(x_3 - h + \alpha \cdot x_1)^2}{2 \cdot \sigma_3^2(x_1)}\right]}$$

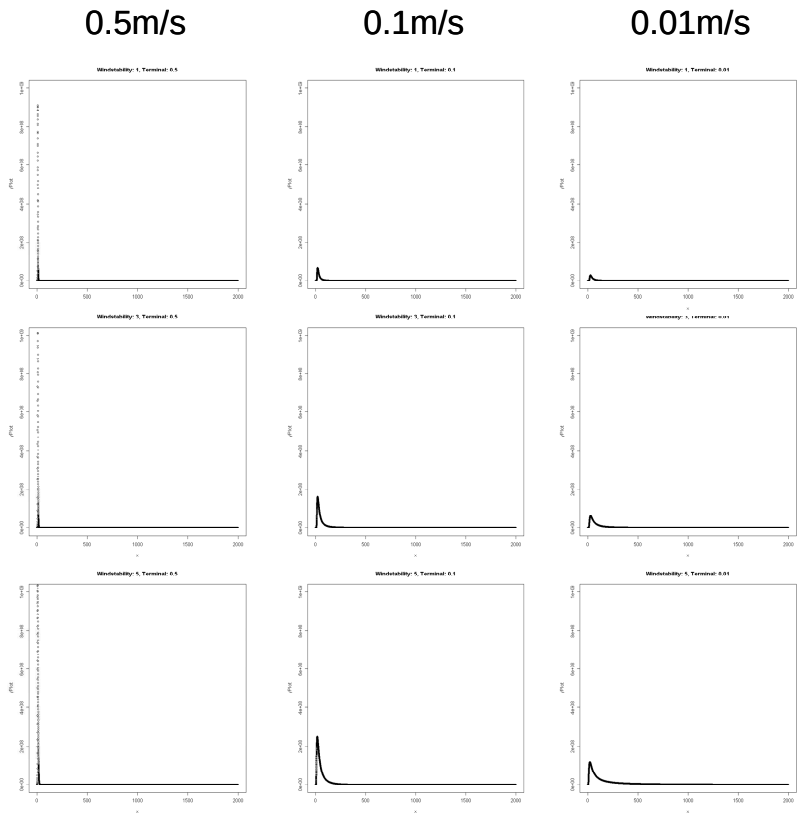
The factor α represents the terminal velocity of poplar pollen v and the constant wind flow in the direction x_1. Therefore α has to be multiplied with x_1. The resulting term responsible for the change of pollen position in vertical direction reduces the difference between the measured height x_3 and the height of the source h.

Equation 4:

$$\alpha = \frac{v}{u}$$

In literature on simulation of dispersal heavy particles have a terminal velocity v of 0.4 m/s whereas light particles have v = 0.04 m/s. The exact terminal velocity of poplar pollen is measured with v = 0.015 m/s and therewith is even lighter. Comparing the results using different terminal velocities and wind stabilities we can see that, as expected, in all cases heavy pollen reaches the ground faster than light pollen. However, it is surprising that we detect more pollen with increasing wind stability (Fig. 6).

(a)



(b)

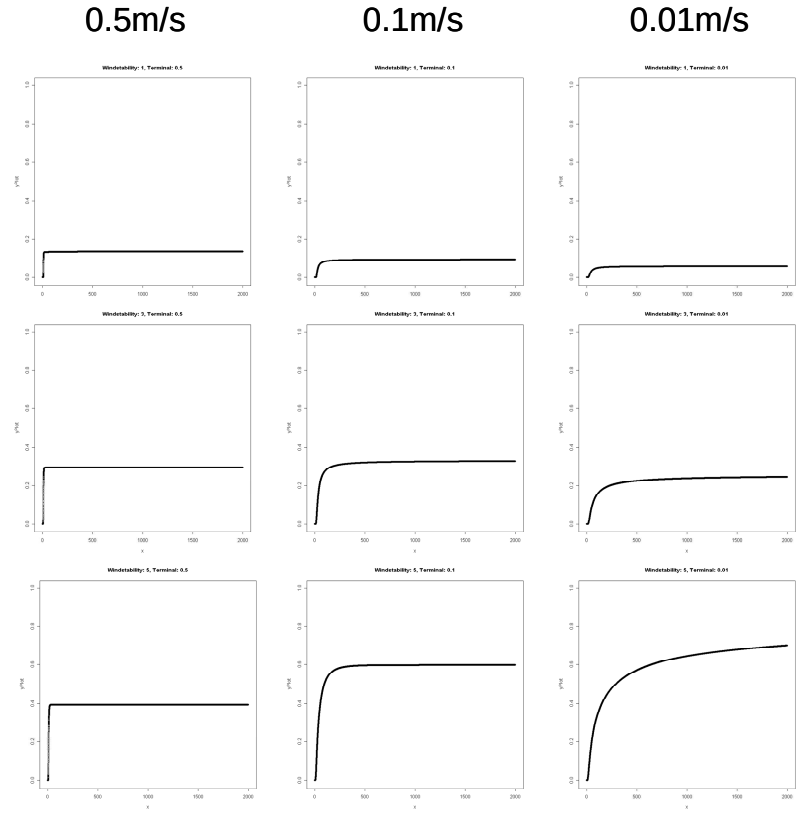


Figure 6: Diagram of the (a) total amount of pollen and (b) cumulative amount of pollen for unstable (first row) to stable (third row) wind conditions

This becomes clear when we do not only count the number of pollen in x_1 but also in x_2 direction (Fig. 7). Because of the responsible σ_2 in equation 3 the pollen is dispersed more in x_2 than in x_1 direction using instable conditions.

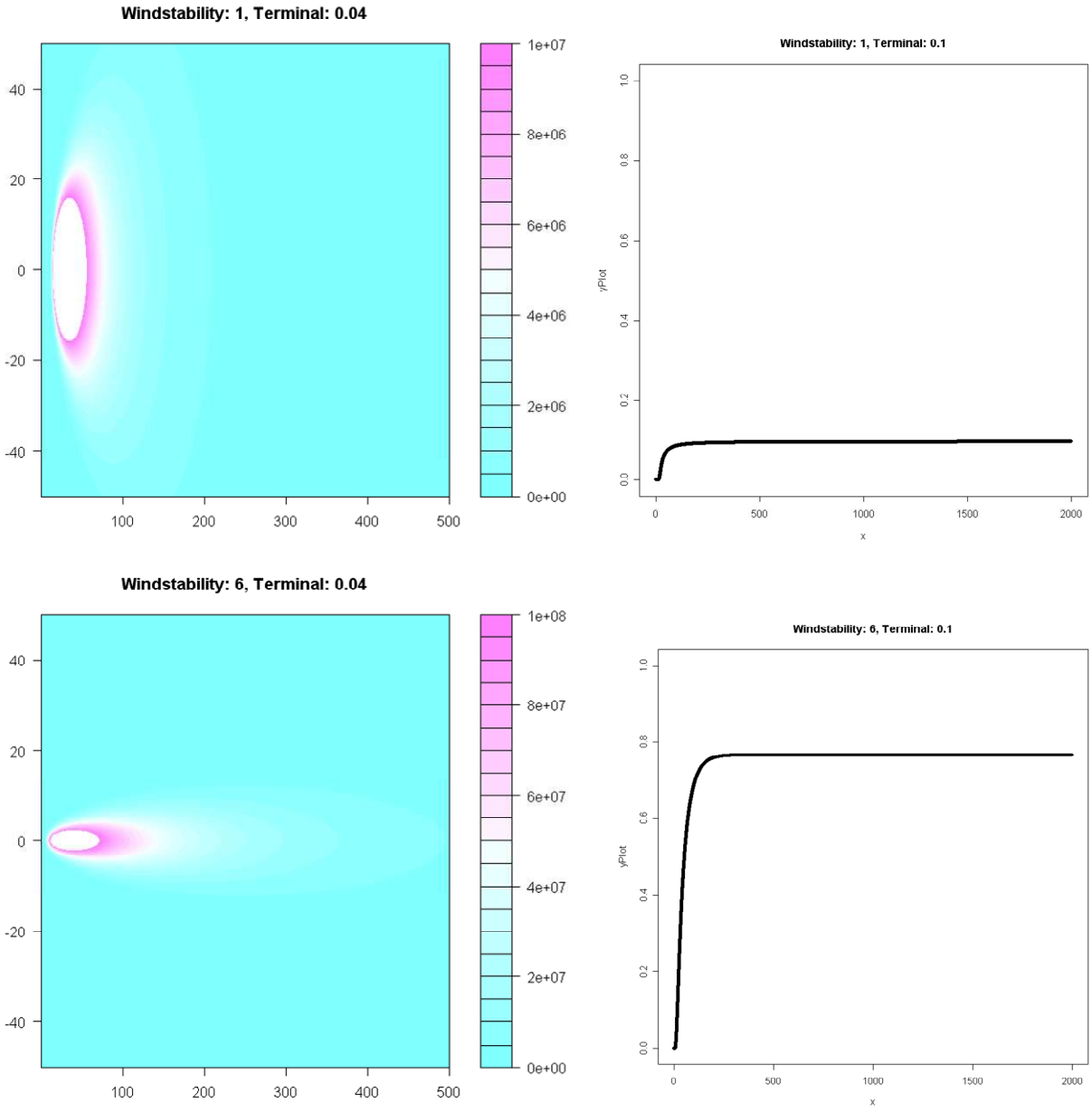


Figure 7: Instable (top) and stable wind conditions (bottom) in x_1 and x_2 direction as well as a cumulative count of pollen for $x_2 = 0$ (right).

Another problem might be the fact that equation 3 as well as equation 1 is not normalised Gaussian distribution function enclosing a total area of 1. Therefore the cumulative amount of pollen even on short distances is able to be bigger than the amount of pollen dispersed from the source. To solve this problem the equation should not be used within a short distance

calculated by equation 5. The distance could reproduce the diameter of the crown of a poplar tree. This would result in a change from a point source to a 2-dimensional source.

Equation 5:

$$y_{Cr} = \frac{\sigma_2(x_0)}{3}$$

When calculating the distance x_0 using the parameters for crown diameter y_{Cr} = 15m, F = 0.306 and f = 0.885 and simulating an unstable condition, we got x_0 = 281.28m.
If this is the range where we cannot use equation 1 and 3, then we will not use the equations at all because the range up to 300m is the most important one for the simulation of pollen dispersal in our study species.

Conclusion

The used wind dependent dispersal model has got tremendous limitations. This based on its origin as a diffusion model for particles under simplified conditions. These limitations are partly due to the simplifications during the Tracer-experiment in 1985 and lead to a dead end in all our attempts to compare the model in theoretical questions with other complex models from the literature. Nevertheless it was possible to simulate pollen dispersal getting realistic results in our study species.
Depending on the characteristic of the local wind directions and the different wind stabilities during pollen release the function should be recognised as an alternative for the simple circular dispersal kernels when modelling pollen dispersal in wind pollinated species.

References

Austerlitz, F., Dick, C.W., Dutech, C., Klein, E.K., Oddou-Muratorio, S., Smouse, P.E. and Sork, V.L. (2004) Using genetic markers to estimate the pollen dispersal curve. Molecular Ecology 13, 937-954.

Kuparinen, A., Markkanen, T., Riikonen, H. and Vesala, T. (2007) Modeling air-mediated dispersal of spores, pollen and seeds in forested areas. Ecological Modelling 208, 177-188.

Overcamp, T.J. (1976) A general Gaussian diffusion-deposition model for elevated point sources. Journal of Applied Meteorology 15, 1167-1171.

Appendix B

Flower phenology of *Populus nigra* and *Populus x canadensis*

Methods

The study site was a natural population of *Populus nigra* located at 51°07'10'' North/ 9°19'44'' South. *Populus x canadensis* trees were scattered within the natural population. 17 female *P. nigra*, 21 male *P. Nigra*, 4 female *P. x canadensis*, 8 male *P. x canadensis* were used in the comparison of the flower phenology. Data collection was conducted one to two times a week between March and June 2006 (17.3.; 28.3.; 2.4.; 10.4.; 19.4.; 25.4.; 2.5.; 11.5.; 17.5.; 7.6.). We defined six stages for the female flowers and four stages for the male flowers based on the phenological stages by M. Villar, INRA, France (personal communication).

The comparison of the most important stage 3 (male pollen release and female receptivity) between the two poplar species was done with the following data:

For every observation time the number of trees in stage 3 were counted. A possible column for male *P. nigra* trees could look like: 90,99,99,110,110,110,115 (One tree was in stage 3 on the 90th day of the year, two trees were in stage 3 on the 99th day, three trees were in stage 3 on the 110th day, one tree was in stage 3 on the 115th day).

Problems during the data collection:

1. The precisely classification into the stages was not always possible. a) The trees were too high for a binocular with a 10 fold magnification. b) Because it was our first time studying flower phenology, we had no experience only pictures and a simple manual how the male and female flowers develop and change during the different stages.
2. The study site was too far away and time was always limited, so that we often could not collect the data more than once per week.
3. The trees were situated within a nature protected area in the floodplain of the Eder River. Observations could therefore only be conducted from the footpath away from the trees.
4. The dynamic floodplain prohibited data collection during high flood. Some areas were not accessible by foot because of temporary streams through the study area.
5. As the study area and the number, the sex and the species status of the trees were new for us, it was not possible to choose a good male-female proportion and enough hybrid trees for comparison with the natural *P. nigra* from the beginning.

Results

There were no significant differences in the flowering phenology between *P. nigra* and *P. x canadensis* trees in the floodplain of the Eder River in the year 2006 (Fig. 1). The male *P. nigra* trees released their pollen between the 10th April and the 2nd May (between the 19th and 25th April pollen release was observed). The female *P. nigra* trees were receptive between the 2nd April and 17th May (between the 19th and 11th May receptivity was observed). The male *P. x canadensis* trees released their pollen between the 10th April and the 11th May (between the 19th April and 2nd May pollen release was observed). The female *P. x canadensis* trees were receptive between the 2nd April and 17th May (between the 19th and 11th May receptivity was observed).

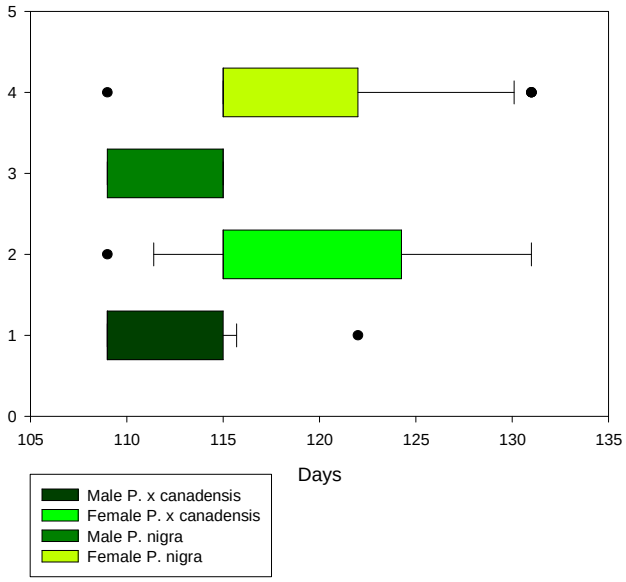
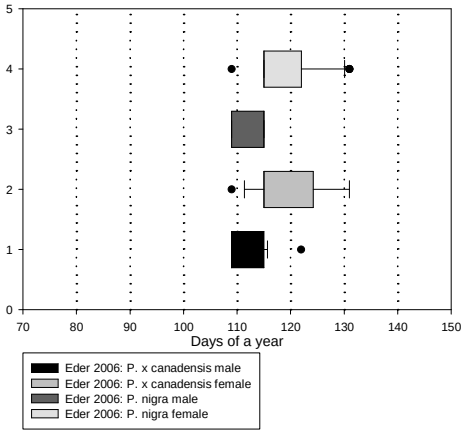
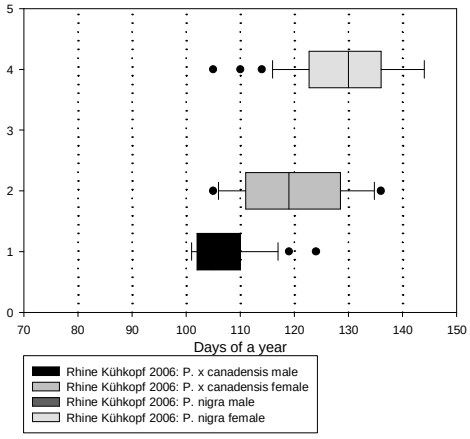


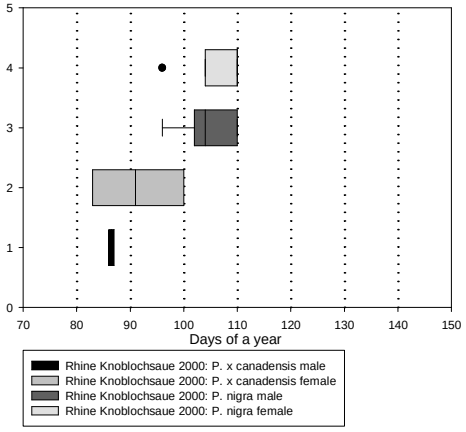
Figure 1: The time period of pollen release and female fecundity in *P. nigra* and *P. x canadensis* in the floodplain of the Eder River, Hesse, Germany in 2006.



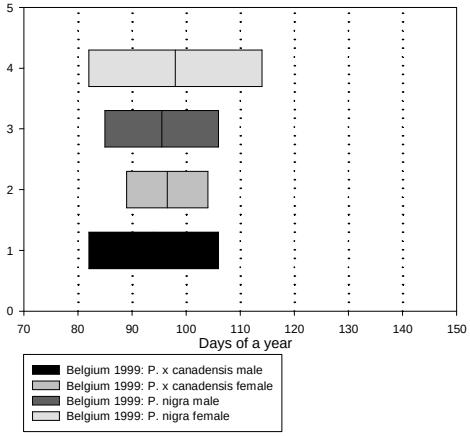
Eder 2006



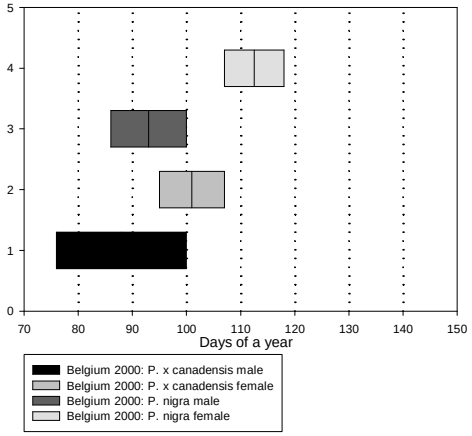
Rhine Kühkopf 2006



Rhine Knoblochsaue 2000 (Gebhardt, unpublished)



Belgium 1999 (vanden Broeck et al. 2003)



Belgium 2000 (vanden Broeck et al. 2003)

Figure 2: Flower phenology of male and female *Populus nigra* and *P. x canadensis* trees at (a) the Eder river in 2006, (b) the Kühkopf 2006, (c) the Knoblochsaue near Kühkopf 2000, (d) the Meuse river in Belgium in 1999 and (e) the Meuse river in Belgium 2000

Conclusion

Comparing our results with the literature and personal communications there were huge differences between regions and years in the start of flowering (Fig. 2). It is known that temperature is the main influencing factor for the start and duration of the flowering period in temperate trees. A combination of chilling and frost in autumn and winter (Partanen et al. 1998) and on the other site the sum of warm temperatures in spring (Fitter et al. 1995, Diekmann 1996) affect the flower phenology. Because these climate conditions vary between regions and years, the analysed differences in our trees and literature mirror these differences in place and time.

References

Diekmann, M. (1996) Relationship between flowering phenology of perennial herbs and meteorological data in deciduous forests of Sweden. Canadian Journal of Botany 74, 528-537.

Fitter, A.H., Fitter, R.S.R., Harris, I.T.B. and Williamson, M.H. (1995) Relationships between 1st flowering date and temperature in the flora of a locality in Central England. Functional Ecology 9, 55-60.

Partanen, J., Koski, V. and Hanninen, H. (1998) Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (*Picea abies*). Tree Physiology 18, 811-816.

Vanden Broeck, A., Cox, K., Quataert, P., Van Bockstaele, E. and Van Slycken, J. (2003) Flowering phenology of *Populus nigra* L., *P. nigra cv. italica* and *P. x canadensis* Moench. and the potential for natural hybridisation in Belgium. Silvae Genetica 52, 280-283.

Appendix C

Germination experiments of *Populus nigra* and *Populus x canadensis*

Methods

Seeds of *Populus nigra* and *Populus x canadensis* trees were harvested in the study area on two days in 2006 and on three days in 2007 due to different stages of the trees with a ”Ruthmann Steiger” and a pruning shear with a telescope shaft. Seed capsules were harvested when the wool was partly visible and some capsules on the branch were still closed. The seed capsules were dried so that all seeds were released with their wool and the wool was totally extended. The seeds were released from the wool by a gadget working with a vacuum cleaner. The Petri-dishes were filled with Vermiculite and filter paper. 50 seeds of each of the 16 mother trees in 2006 (*P. nigra*: 1, 7, 9, 17, 108, 169, 205, 218, 228; *P. x canadensis*: 29, 42, 80, 82, 83, 146, 152) and of the 13 mother trees in 2007 (*P. nigra*: 1, 9, 17, 169, 216, 228, 2001.1; *P. x canadensis*: 80, 83, 146, 152, 1156.6, 2025.2) were randomly picked and were put into each of four Petri-dishes per tree (total number of Petri-dishes 2006: 64; 2007: 52). The Petri-dishes were kept moistly for the whole experiment. The germinated seeds were counted after 4 days, harvested and transferred to the fridge for further DNA studies.

Results

In 2006 there is a significant difference between the germination rates of seeds from a female *P. nigra* and *P. x canadensis* tree ($p<0.05$). The mean total germination rate for all trees of one species is slightly higher for *P. nigra* than for *P. x canadensis* (0.65 +/- 0.21 vs. 0.54 +/- 0.25; see also Fig. 1).

In 2007 there is again a significant difference between the two species ($p<0.05$). The mean total germination rate for all trees of one species is for both species close to 100% (*P. nigra*: 0.93 +/- 0.06 vs. *P. x canadensis*: 0.89 +/- 0.09; see also Fig. 2).

As there is a highly significant difference between the years for the germination rates, the datasets were not pooled.

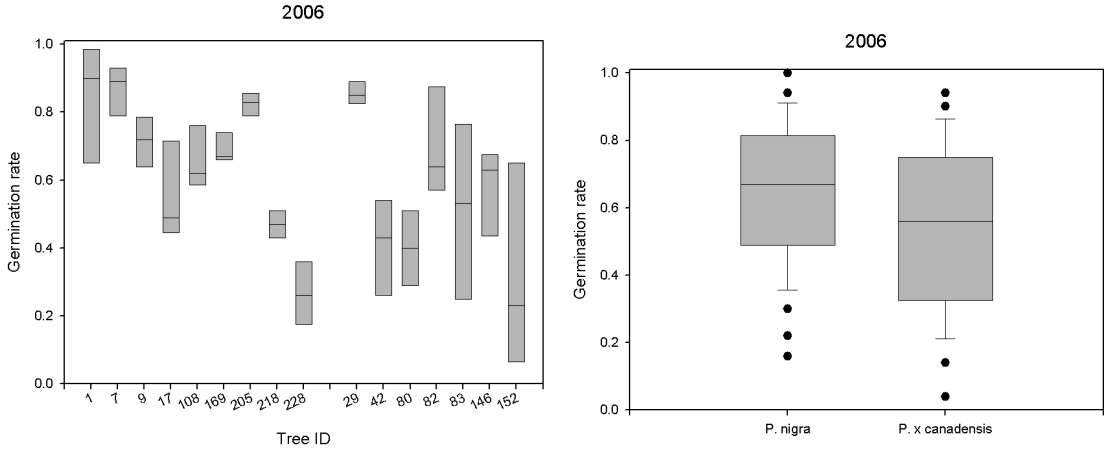


Figure 1: Germination rates of the studied trees differentiated in *P. nigra* (left bars) and *P. x canadensis* mothers (right bars) in 2006

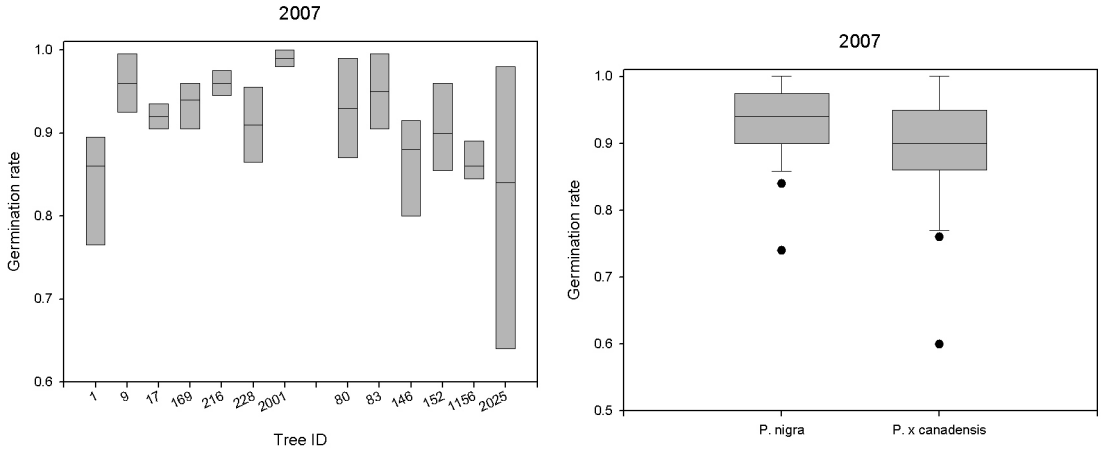


Figure 2: Germination rates of the studied trees differentiated in *P. nigra* (left bars) and *P. x canadensis* mothers (right bars) in 2007

Discussion

The differences between the years based on the different times of harvesting. In 2006, some capsules might not been totally ripened, so that the germination rate is lower than in 2007. Because *P. nigra* is a pioneer species the germination rates are usually high (Karrenberg and Suter 2003).

It has to be mentioned that the seeds were harvested directly from the tree in the study region and were not hand pollinated. Therefore, the germination rates and the differentiation into *P. nigra and P. x canadensis* mirror only the female proportion in the seed. The father trees could be *P. nigra* as well as *P. x canadensis* trees. The results and the differentiation might be clearer if hand pollination experiments are conducted.

References

Karrenberg, S. and Sutter, M. (2003) Phenotypic trade-offs in the sexual reproduction of Salicaceae from flood plains. American Journal of Botany 90, 749-754.

SUMMARY

DISPERSAL IS AN ESSENTIAL PROCESS IN PLANT SURVIVAL because it allows plants as static organism to move. Plant dispersal is necessary to avoid unsuitable conditions and to colonise new sites. Additionally gene flow between populations reduces possible genetic threads as inbreeding depression. Genetic exchange is not limited to the dispersal of seeds, but is also possible by pollen dispersal.

The most common dispersal vectors for plants are animals and wind. However, in human dominated landscapes animals could be replaced by humans. To assess the importance of humans on plant distribution a simulation model was created. The effect of human movement and gardening styles were tested for different ruderal plant species. I found strong influence of human movement behaviour especially on alien species. During 20 years these species extended their distribution range taking benefit from increased human mobility as revealed by sociological studies. In contrast native species tended to be more affected by the change of the agricultural landscape and of rural to urban life styles. Nevertheless, the change in species distribution could in all cases be better explained by simulating human transport based on human movement behaviour than by a distance dependent diffusion model.

Another approach using simulation models for dispersal has been done in the wind pollinated floodplain species *Populus nigra* in Central Germany. The aim of the study was to assess the extent of intraspecies respectively introgressive pollen mediated gene flow by the hybrid *P. x canadensis*. Therefore the estimation of a pollen dispersal kernel was essential to model the dispersal processes. Genetic paternity analyses allow the calculation of pollen dispersal distances between father and mother trees. I used marked point pattern analyses to estimate and assess the uncertainty in fitted pollen dispersal kernels. There was only a significant departure from the null random mating model up to a distance of approximately 300m, although we were dealing with a wind pollinated species and the analyses were based on a huge dataset of more than 1,500 seeds. A two-component pollen dispersal kernel comprising an exponential power function and a truncated uniform function was the most parsimonious model to fit the data. The fitted kernel was able to predict the number of seeds fathered by a male tree and provided comparable results with published spatial statistic models. Our results strongly suggest that kernel estimates based on the direct fit of the observed mating distances differed considerably from fitted kernels that account for the spatial structure of adult trees. The uncertainty in the estimation of those directly fitted kernel may be underestimated.

With a suitable pollen dispersal kernel at hand, simulation studies were conducted to understand the discrepancy between low introgression rates in the natural population of *P. nigra* and the overwhelming number of hybrids in the close vicinity. Using different

standard dispersal kernels and our fitted two-component exponential power function, we found that the probability of fertilisation was 10^{-1} to 10^{-2} times lower in hybrid than in *P. nigra* pollen. This range could be confirmed with published results from hand pollination experiments and is the first time that the reproductive barrier between *P. nigra* and *P. x canadensis* has been revealed in open pollinated trees in a natural population.

Using simulation models to understand different kinds of dispersal mechanisms turned out to be a promising way not only to describe the process of dispersal itself but also to combine expert knowledge of various disciplines in an ecological simulation model.

ZUSAMMENFASSUNG – GERMAN SUMMARY

AUSBREITUNG IST EIN ENTSCHEIDENDER PROZESS IM ÜBERLEBEN VON PFLANZEN, denn er ermöglicht Pflanzen als sessilen Organismen sich fortzubewegen. Ausbreitung ist sowohl zur Vermeidung von schlechten Bedingungen als auch zur Neubesiedlung von Flächen notwendig. Außerdem reduziert ein genetischer Austausch zwischen Populationen mögliche genetische Bedrohungen wie z.B. Inzuchtdepression.

Die häufigsten Ausbreitungsvektoren bei Pflanzen sind Tiere und Wind. In vom Menschen dominierten Landschaften jedoch können Menschen die tierische Verbreitung ersetzen. Um die Bedeutung des Menschen auf die Pflanzenverbreitung zu untersuchen wurde ein Simulationsmodell erstellt. Der Effekt der menschlichen Bewegung und Gartenstile wurde für verschiedene ruderale Pflanzen getestet. Ich fand einen starken Einfluss des menschlichen Bewegungsverhaltens besonders bei nicht-einheimischen Arten heraus. Innerhalb von 20 Jahren vergrößerten diese Arten ihre Verbreitung, wobei sie von der ansteigenden Mobilität der Menschen profitierten, die in soziologischen Untersuchungen nachgewiesen worden war. Im Gegensatz dazu schienen einheimische Arten stärker auf den Wechsel der bäuerlichen Landschaft und von einem ländlichem zu einem städtischen Lebensstil zu reagieren. Allerdings konnte die Veränderung in der Artverbreitung in allen Fällen besser mittels einer menschlichen Ausbreitung basierend auf menschlichen Bewegungsverhalten simuliert werden als durch ein distanzabhängiges Diffusionsmodell.

Eine weitere Herangehensweise zur Nutzung von Simulationsmodellen bei Ausbreitungsprozessen wurde anhand der windbestäubten, in der Weichholzaue von Flüssen vorkommenden Art *Populus nigra* in Mitteldeutschland durchgeführt. Das Ziel der Untersuchung war es, das Ausmaß von innerartlichem Pollenvermittelten Genfluß durch den Schwarzpappelhybriden *P. x canadensis* abzuschätzen.

Daher war die Abschätzung eines passenden Ausbreitungskernel entscheidend für die Modellierung der Ausbreitung. Genetische Vaterschaftsanalysen ermöglichen die Berechnung von Ausbreitungsdistanzen des Pollens zwischen Vater- und Mutterbäumen. Ich habe Punktmusteranalysen benutzt um Ausbreitungskernel anzupassen und deren statistische Unsicherheit zu testen. Es gab nur bis zu einer Distanz von 300m eine signifikante Abweichung vom Paarungs-Nullmodell, obwohl es sich um eine windbestäubte Art handelt und die Analysen auf einem sehr großen Datensatz von mehr als 1.500 Samen beruhten. Ein Zwei-Komponenten-Ausbreitungskernel aus einer exponential-power Funktion und einer eingeschränkten Gleichverteilungsfunktion bildete das am besten angepasste Model zu den Daten. Der angepasste Kernel konnte die Anzahl von Samen, die von einem männlichen Baum bestäubt wurden, abschätzen und lieferte vergleichbare Ergebnisse wie veröffentlichte

räumlich statistische Modelle. Unsere Ergebnisse zeigen deutlich, dass Ausbreitungskernel, die auf einer direkten Anpassung an die beobachteten Paarungsdistanzen beruhen, stark von Ergebnissen abweichen, die durch Methoden erzielt wurden, die die räumliche Struktur von Altbäumen berücksichtigen. Die Unsicherheit in der Abschätzung solcher direkt angepasster Ausbreitungskernel scheint unterschätzt zu werden.

Mit dem angepassten Ausbreitungskernel wurden Simulationen durchgeführt, um die Diskrepanz zwischen geringer Einkreuzung durch den Hybriden in der natürlichen Schwarzpappelpopulation und der überwältigenden Anzahl dieser Hybride in der direkten Nachbarschaft zu untersuchten. Durch die Verwendung von verschiedenen standardmäßigen Ausbreitungskernel und des angepassten Zwei-Komponenten-Ausbreitungskernel fanden wir heraus, dass die Wahrscheinlichkeit einer Befruchtung durch Hybridpollen 10^{-1} bis 10^{-2} mal geringer als bei der Schwarzpappel war. Dieser Werteraum konnte über Ergebnisse bestätigt werden, die in Handbestäubungsexperimenten entstanden. Dies ist das erste Mal, dass die reproduktive Barriere zwischen *P. nigra* und *P. x canadensis* in frei bestäubten Bäumen in einem natürlichen Bestand gezeigt werden konnte.

Die Verwendung von Simulationsmodellen zur Untersuchung von unterschiedlichen Ausbreitungsmechanismen stellte sich als viel versprechend heraus, nicht nur zur Beschreibung des Ausbreitungsprozesses selbst, sondern auch hinsichtlich der Kombination von Expertenwissen aus den verschiedensten Fachrichtungen in einem ökologischen Modell.

ACKNOWLEDGEMENTS

Ich danke Birgit Ziegenhagen, in deren Arbeitsgruppe ich meine Promotionsarbeit durchführen konnte, für ihre Betreuung und die Unterstützung, auf die man sich immer verlassen konnte, sowie für die familiäre Atmosphäre, in der ich mich sehr wohl gefühlt habe.

Mein ganz besonderer Dank gilt Ronald Bialozyt, meinem Betreuer, Zimmergenossen und gutem Freund für die vielen Gespräche über alles Relevante und Irrelevante. Es war eine schöne Zeit.

Danken möchte ich der gesamten Arbeitsgruppe Naturschutzbiologie, die während der letzten Jahre mein Leben bestimmt hat, angenehmes Arbeitsumfeld war und an die ich gerne zurück denken werde.

Ich danke Volker Grimm für das Zweitgutachten und besonders dafür, dass er mit seiner Vorlesung „Theoretische Ökologie“ die Weichen für mein Interesse an Modellen geweckt hat und mir im Weiteren wissenschaftlich stets zur Seite stand.

Ich danke den weiteren Mitgliedern der Prüfungskommission Roland Brandl und Joachim Schachtner für ihren Beitrag zum erfolgreichen Abschluss meiner Promotion.

Weiterhin danke ich den Mitgliedern des Forschungszentrums für Gesellschaft und Ökologie Stefan Brunzel und Jens Jetzkowitz für die Zusammenarbeit an meinem ersten Modell.

Ich danke James Bullock für die Einladung nach England und die Unterstützung meiner Arbeit. Ebenfalls danke ich Matthias Wichmann und den drei Piraten für die wunderschöne Zeit in Dorset, die ich niemals vergessen werde.

Danken möchte ich Jürgen Vollmer für die Stunden mit Zettel und Stift zur Umsetzung von Mathematik und Physik in einem biologischen Zusammenhang, Bernhard Vogel für Ideen und Diskussionen zur Windausbreitung und den Mitgliedern der OESA Leipzig für viele kleine und große aber immer fruchtbare Diskussionen bezüglich meiner Arbeit.

Ich danke dem BMBF für die finanzielle Unterstützung von Teilen der Arbeit sowie dem europäischen Forschungsnetzwerk EVOLTREE.

Ein besonderer Dank geht an alle meine Freunde für die gemeinsame Zeit, speziell Matze und Steffi für abwechslungsreiche Mensagänge und lange Spielabende, und Matthias für ein offenes Ohr bei den unterschiedlichsten wissenschaftlichen und unwissenschaftlichen Problemen.

Mein abschließender Dank gilt meiner Familie für die immerwährende Unterstützung in jeglicher Form und die entspannten Wochenenden daheim.

Thanks to everyone supporting me and my work.

Curriculum vitae

MARC NIGGEMANN

Current address

Conservation Biology
Karl-von-Frisch-Straße 8
35032 Marburg/ Germany
Phone: +49 6421 2824878
Email: marc.niggemann@staff.uni-marburg.de

Personal

Date and place of birth:	30.07.1977 in Arnsberg/ Germany
Marital status:	Single

Education

2004 – present	Doctoral thesis, Philipps-University Marburg/ Germany
2004.04.	Diploma, Philipps-University Marburg/ Germany; Title of diploma thesis: The effect of population size and density on the reproduction of *Trifolium montanum* L. and on the behaviour of its pollinators.
1998 – 2004	Undergraduate, Philipps-University Marburg/ Germany
1988 – 1997	High School, Gymnasium Laurentianum Arnsberg/ Germany

Further education

2008.11.	"Using Geodatabases in ArcGis" by ESRI, Germany
2004.04.	"Ecological Modelling" by the UFZ, Leipzig/ Germany

Academic appointment

2005 – 2008	Research assistant, Philipps-University Marburg/ Germany
2004 – 2005	Graduate assistant, Philipps-University Marburg/ Germany

Fellowship

2007.02. - 2007.04.	10-week stay in the CEH Dorset/ UK funded by the European Network of Excellence Evoltree

Teaching

Mentor for Bachelor and Diploma students
Tudor in Undergraduate courses:

- Modelling in conservation biology
- System analyses and simulation
- Using Geographic Information Systems

Tutor in the 9th UFZ Winterschule for Ecological Modelling in Leipzig

Publications

Reviewed articles:

- Rathmacher, G., Niggemann, M., Köhnen, M., Ziegenhagen, B. & R. Bialozyt. Short-distance gene flow in of *Populus nigra* L. accounts for small scale spatial genetic structures – implications for in situ conservation measures, Conservation Genetics, accepted.
- Rathmacher, G., Niggemann, M., Wypukol, H., Gebhardt, K., Ziegenhagen, B., & R. Bialozyt (2009). Allelic ladders and reference genotypes for a rigorous standardization of poplar microsatellite data. TREES – Structure and Function 23: 573-583.
- Niggemann, M., Jetzkowitz, J., Brunzel, S., Wichmann, M.C. & R. Bialozyt (2009). Distribution patterns of plants explained by human movement behaviour. Ecological Modelling, 220: 1339-1346.
- Wichmann, M.C., Alexander, M.J., Soons, M.B., Galsworthy, S., Dunne, L., Gould, R., Fairfax, C., Niggemann, M., Hails, R. & J. M. Bullock (2009). Human mediated dispersal of seeds over long distances. Proceedings of the Royal Society B, 276: 523-532.
- Schleuning, M., Niggemann, M., Becker, U. & D. Matthies (2009). Negative effects of habitat degradation and fragmentation on the declining grassland plant *Trifolium montanum*. Basic and Applied Ecology, 10: 61-69.

Submitted:

- Niggemann, M., Wiegand, T., Robledo-Arnuncio, J.J. & R. Bialozyt. Marked point pattern analysis on genetic paternity data for estimation and uncertainty assessment of pollen dispersal kernels. Journal of Ecology, submitted.

Abstracts:

- Flinkerbusch, S., Heymann, E., Niggemann, M. & R. Bialozyt (2008). Simulation of seed dispersal by two neo tropical tamarin species. -, in: Verhandlungen der Gesellschaft für Ökologie (GfÖ), Band 38: 38th Annual conference of the GfÖ in Leipzig.
- Niggemann, M., Rathmacher, G., & R. Bialozyt (2008). A new approach for modelling wind dispersal. 4th European Symposium on Aerobiology in Turku, Finland.

- Niggemann, M. & R. Bialozyt (2007). How complex should a dispersal kernel be? Comparing dispersal kernels on their usability for modelling dispersal events, in: Conference proceedings of the 6. European Conference on Ecological Modelling in Trieste, Italy.
- Niggemann, M., Jetzkowitz, J., Brunzel, S., Wichmann, M.C. & R. Bialozyt (2007). Modelling human-mediated dispersal in ruderal plants using biological and sociological data, in: Conference proceedings of the 6. European Conference on Ecological Modelling in Trieste, Italy.
- Niggemann, M., Rathmacher, G. & R. Bialozyt (2007). The risk of introgression of foreign genes in Populus spec. - Does a complex pollen dispersal kernel always lead to a better fit of the model -, in: Verhandlungen der Gesellschaft für Ökologie (GfÖ), Band 37: 37th Annual conference of the GfÖ in Marburg.
- Niggemann, M., Rathmacher, G. & R. Bialozyt (2006). The risk of introgression of foreign genes in Populus spec. - Differences in the Flowering Phenology of *Populus nigra* and *Populus x canadensis* -, in: Verhandlungen der Gesellschaft für Ökologie (GfÖ), Band 36: 36th Annual conference of the GfÖ in Bremen.
- Niggemann, M., Rathmacher G. & R. Bialozyt (2006). The risk of introgression of foreign genes in Populus spec. - Modelling introgressive gene flow in the landscape -, in: Abstract book of the 1. European Congress of Conservation Biology in Eger, Hungary.
- Niggemann, M., Fiedler, T. & R. Bialozyt (2005): Estimating the dispersal abilities of ruderal plant species: Modelling sociological and biological data, in: Verhandlungen der Gesellschaft für Ökologie (GfÖ), Band 35: 35th Annual conference of the GfÖ in Regensburg.
- Niggemann, M. & R. Bialozyt (2005). Human induced seed dispersal - Modelling sociological and biological data, in: Contributions to the 18th Annual Conference of the GfÖ - Section Plant Population Biology in Potsdam.
- Niggemann, M. & D. Matthies (2004): The effect of population size and density on the reproduction of *Trifolium montanum* L., in: Verhandlungen der Gesellschaft für Ökologie (GfÖ), Band 34: 34th Annual conference of the GfÖ in Gießen.
- Schleuning, M., Niggemann, M. & D. Matthies (2003): Quantitative genetic variation in *Trifolium montanum*, an endangered plant of calcareous grasslands, in: Verhandlungen der Gesellschaft für Ökologie (GfÖ), Band 33: 33rd Annual conference of the GfÖ in Halle (Poster award).

Marburg, July 2009

Marc Niggemann

Ich versichere, dass ich meine Dissertation

Modelling dispersal in plants (Modellierung der Ausbreitung bei Pflanzen)

selbstständig, ohne unerlaubte Hilfe angefertigt und mich dabei keiner anderen als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe.

Die Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg 2009

(Marc Niggemann)